





# Advanced Analysis

Min Yan

Department of Mathematics

Hong Kong University of Science and Technology

December 19, 2018

# Contents

<b>1</b>	<b>Limit of Sequence</b>	<b>7</b>
1.1	Definition . . . . .	8
1.2	Property of Limit . . . . .	13
1.3	Infinity and Infinitesimal . . . . .	17
1.4	Supremum and Infimum . . . . .	19
1.5	Convergent Subsequence . . . . .	24
1.6	Additional Exercise . . . . .	33
<b>2</b>	<b>Limit of Function</b>	<b>37</b>
2.1	Definition . . . . .	38
2.2	Basic Limit . . . . .	46
2.3	Continuity . . . . .	53
2.4	Compactness Property . . . . .	58
2.5	Connectedness Property . . . . .	62
2.6	Additional Exercise . . . . .	70
<b>3</b>	<b>Differentiation</b>	<b>77</b>
3.1	Linear Approximation . . . . .	78
3.2	Computation . . . . .	87
3.3	Mean Value Theorem . . . . .	94
3.4	High Order Approximation . . . . .	105
3.5	Application . . . . .	114
3.6	Additional Exercise . . . . .	119
<b>4</b>	<b>Integration</b>	<b>125</b>
4.1	Riemann Integration . . . . .	126
4.2	Darboux Integration . . . . .	135
4.3	Property of Riemann Integration . . . . .	142
4.4	Fundamental Theorem of Calculus . . . . .	149
4.5	Riemann-Stieltjes Integration . . . . .	155
4.6	Bounded Variation Function . . . . .	163
4.7	Additional Exercise . . . . .	172
<b>5</b>	<b>Topics in Analysis</b>	<b>179</b>

---

5.1	Improper Integration . . . . .	180
5.2	Series of Numbers . . . . .	184
5.3	Uniform Convergence . . . . .	196
5.4	Exchange of Limits . . . . .	207
5.5	Additional Exercise . . . . .	217
<b>6</b>	<b>Multivariable Function</b>	<b>225</b>
6.1	Limit in Euclidean Space . . . . .	226
6.2	Multivariable Map . . . . .	233
6.3	Compact Subset . . . . .	242
6.4	Open Subset . . . . .	248
6.5	Additional Exercise . . . . .	254
<b>7</b>	<b>Multivariable Algebra</b>	<b>257</b>
7.1	Linear Transform . . . . .	258
7.2	Bilinear Map . . . . .	264
7.3	Multilinear Map . . . . .	274
7.4	Orientation . . . . .	284
7.5	Additional Exercises . . . . .	291
<b>8</b>	<b>Multivariable Differentiation</b>	<b>293</b>
8.1	Linear Approximation . . . . .	294
8.2	Property of Linear Approximation . . . . .	303
8.3	Inverse and Implicit Differentiations . . . . .	309
8.4	Submanifold . . . . .	317
8.5	High Order Approximation . . . . .	325
8.6	Maximum and Minimum . . . . .	334
8.7	Additional Exercise . . . . .	346
<b>9</b>	<b>Measure</b>	<b>351</b>
9.1	Length in $\mathbb{R}$ . . . . .	352
9.2	Lebesgue Measure in $\mathbb{R}$ . . . . .	358
9.3	Outer Measure . . . . .	362
9.4	Measure Space . . . . .	368
9.5	Additional Exercise . . . . .	376
<b>10</b>	<b>Lebesgue Integration</b>	<b>379</b>
10.1	Integration in Bounded Case . . . . .	380
10.2	Measurable Function . . . . .	385
10.3	Integration in Unbounded Case . . . . .	392
10.4	Convergence Theorem . . . . .	404
10.5	Convergence and Approximation . . . . .	411
10.6	Additional Exercise . . . . .	416
<b>11</b>	<b>Product Measure</b>	<b>419</b>
11.1	Extension Theorem . . . . .	420

---

11.2	Lebesgue-Stieltjes Measure . . . . .	427
11.3	Product Measure . . . . .	433
11.4	Lebesgue Measure on $\mathbb{R}^n$ . . . . .	440
11.5	Riemann Integration on $\mathbb{R}^n$ . . . . .	451
11.6	Additional Exercise . . . . .	462
<b>12</b>	<b>Differentiation of Measure</b>	<b>465</b>
12.1	Radon-Nikodym Theorem . . . . .	466
12.2	Lebesgue Differentiation Theorem . . . . .	476
12.3	Differentiation on $\mathbb{R}$ : Fundamental Theorem . . . . .	481
12.4	Differentiation on $\mathbb{R}^n$ : Change of Variable . . . . .	491
12.5	Additional Exercise . . . . .	499
<b>13</b>	<b>Multivariable Integration</b>	<b>501</b>
13.1	Curve . . . . .	502
13.2	Surface . . . . .	512
13.3	Submanifold . . . . .	524
13.4	Green's Theorem . . . . .	530
13.5	Stokes' Theorem . . . . .	541
13.6	Gauss' Theorem . . . . .	548
13.7	Additional Exercise . . . . .	553
<b>14</b>	<b>Manifold</b>	<b>555</b>
14.1	Manifold . . . . .	556
14.2	Topology of Manifold . . . . .	564
14.3	Tangent and Cotangent . . . . .	571
14.4	Differentiable Map . . . . .	580
14.5	Orientation . . . . .	588
<b>15</b>	<b>Field on Manifold</b>	<b>597</b>
15.1	Tangent Field . . . . .	598
15.2	Differential Form . . . . .	602
15.3	Lie Derivative . . . . .	609
15.4	Integration . . . . .	615
15.5	Homotopy . . . . .	624
15.6	deRham Cohomology . . . . .	632
15.7	Singular Homology . . . . .	632
15.8	Poincaré Duality . . . . .	632

## Chapter 1

# Limit of Sequence

## 1.1 Definition

A *sequence* is an infinite list

$$x_1, x_2, x_3, \dots, x_n, x_{n+1}, \dots$$

We also denote the sequence by  $\{x_n\}$  or simply  $x_n$ . The subscript  $n$  is the *index* and does not have to start from 1. For example,

$$x_5, x_6, x_7, \dots, x_n, x_{n+1}, \dots,$$

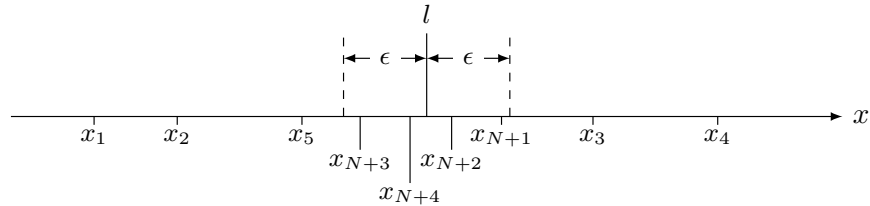
is also a sequence, with the index starting from 5.

In this chapter, the terms  $x_n$  of a sequence are assumed to be real numbers and can be plotted on the real number line.

**Definition 1.1.1.** A sequence  $x_n$  of real numbers has *limit*  $l$  (or *converges* to  $l$ ), and denoted  $\lim_{n \rightarrow \infty} x_n = l$ , if for any  $\epsilon > 0$ , there is  $N$ , such that

$$n > N \implies |x_n - l| < \epsilon. \quad (1.1.1)$$

A sequence is *convergent* if it has a (finite) limit. Otherwise, the sequence is *divergent*.



**Figure 1.1.1.** For any  $\epsilon$ , there is  $N$ .

Since the limit is about the long term behavior of a sequence getting closer and closer to a target, only small  $\epsilon$  and big  $N$  need to be considered in establishing a limit. For example, the limit of a sequence is not changed if the first one hundred terms are replaced by other arbitrary numbers. See Exercise 1.6. Exercise 1.8 contains more examples.

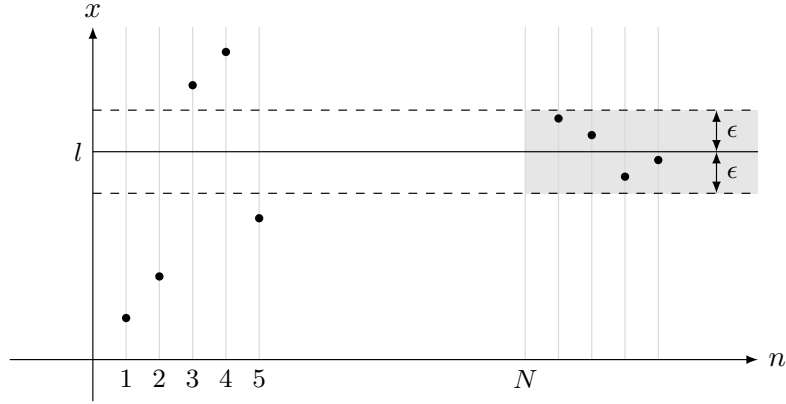
Attention needs to be paid to the logical relation between  $\epsilon$  and  $N$ . The smallness  $\epsilon$  for  $|x_n - l|$  is *arbitrarily* given, while the size  $N$  for  $n$  is to be found *after*  $\epsilon$  is given. Thus the choice of  $N$  usually depends on  $\epsilon$ .

In the following examples, we establish the most important basic limits. For any given  $\epsilon$ , the analysis leading to the suitable choice of  $N$  will be given. It is left to the reader to write down the rigorous formal argument.

**Example 1.1.1.** We have

$$\lim_{n \rightarrow \infty} \frac{1}{n^p} = 0 \text{ for } p > 0. \quad (1.1.2)$$





**Figure 1.1.2.** *Plotting of a convergent sequence.*

Another way of expressing the same limit is

$$\lim_{n \rightarrow \infty} n^p = 0 \text{ for } p < 0.$$

To establish the limit (1.1.2), we note that the inequality  $\left| \frac{1}{n^p} - 0 \right| = \left| \frac{1}{n^p} \right| < \epsilon$  is the same as  $\frac{1}{n} < \epsilon^{\frac{1}{p}}$ , or  $n > \epsilon^{-\frac{1}{p}}$ . Therefore choosing  $N = \epsilon^{-\frac{1}{p}}$  should make the implication (1.1.1) hold.

**Example 1.1.2.** We have

$$\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1. \quad (1.1.3)$$

Let  $x_n = \sqrt[n]{n} - 1$ . Then  $x_n > 0$  and

$$n = (1 + x_n)^n = 1 + nx_n + \frac{n(n-1)}{2}x_n^2 + \dots > \frac{n(n-1)}{2}x_n^2.$$

This implies  $x_n^2 < \frac{2}{n-1}$ . In order to get  $|\sqrt[n]{n} - 1| = x_n < \epsilon$ , it is sufficient to have  $\frac{2}{n-1} < \epsilon^2$ , which is the same as  $N > \frac{2}{\epsilon^2} + 1$ . Therefore we may choose  $N = \frac{2}{\epsilon^2} + 1$ .

**Example 1.1.3.** We have

$$\lim_{n \rightarrow \infty} a^n = 0 \text{ for } |a| < 1. \quad (1.1.4)$$

Another way of expressing the same limit is

$$\lim_{n \rightarrow \infty} \frac{1}{a^n} = 0 \text{ for } |a| > 1.$$

Let  $\frac{1}{|a|} = 1 + b$ . Then  $b > 0$  and

$$\frac{1}{|a^n|} = (1 + b)^n = 1 + nb + \frac{n(n-1)}{2}b^2 + \dots > nb.$$

This implies  $|a^n| < \frac{1}{nb}$ . In order to get  $|a^n| < \epsilon$ , it is sufficient to have  $\frac{1}{nb} < \epsilon$ . This suggests us to choose  $N = \frac{1}{b\epsilon}$ .

More generally, the limit (1.1.4) may be extended to

$$\lim_{n \rightarrow \infty} n^p a^n = 0 \text{ for } |a| < 1 \text{ and any } p. \quad (1.1.5)$$

Fix a natural number  $P > p + 1$ . For  $n > 2P$ , we have

$$\begin{aligned} \frac{1}{|a^n|} &= 1 + nb + \frac{n(n-1)}{2}b^2 + \cdots + \frac{n(n-1)\cdots(n-P+1)}{P!}b^P + \cdots \\ &> \frac{n(n-1)\cdots(n-P+1)}{P!}b^P > \frac{\left(\frac{n}{2}\right)^P}{P!}b^P. \end{aligned}$$

This implies

$$|n^p a^n| < \frac{n^P |a^n|}{n} < \frac{2^P P!}{b^P} \frac{1}{n},$$

and suggests us to choose  $N = \max \left\{ 2P, \frac{2^P P!}{2b^P \epsilon} \right\}$ .

**Example 1.1.4.** For any  $a$ , we have

$$\lim_{n \rightarrow \infty} \frac{a^n}{n!} = 0. \quad (1.1.6)$$

Fix a natural number  $P > |a|$ . For  $n > P$ , we have

$$\left| \frac{a^n}{n!} \right| = \frac{|a|^P}{P!} \frac{|a|}{P+1} \frac{|a|}{P+2} \cdots \frac{|a|}{n-1} \frac{|a|}{n} \leq \frac{|a|^P}{P!} \frac{|a|}{n}.$$

In order to get  $\left| \frac{a^n}{n!} \right| < \epsilon$ , we only need to make sure  $\frac{|a|^P}{P!} \frac{|a|}{n} < \epsilon$ . This leads to the choice

$$N = \max \left\{ P, \frac{|a|^{P+1}}{P! \epsilon} \right\}.$$

Any logical argument needs to start from some known facts. Example 1.1.1 assumes the knowledge of the exponential  $a^b$  for any positive real number  $a$  and any real number  $b$ . Example 1.1.3 makes use of binomial expansion, which is the knowledge about the addition and multiplication of real numbers. Moreover, all the arguments involve the knowledge about the comparison of real numbers.

The knowledge about real numbers is the logical foundation of mathematical analysis. In this course, all the proofs are logically derived from the properties about the *arithmetic operations*  $+$ ,  $-$ ,  $\times$ ,  $\div$  and the *order*  $<$ , which should satisfy many usual properties such as the following (the whole list has more than 20 properties).

- *Commutativity*:  $a + b = b + a$ ,  $ab = ba$ .
- *Distributivity*:  $a(b + c) = ab + ac$ .
- *Unit*: There is a special number 1 such that  $1a = a$ .

- *Exclusivity*: One and only one from  $a < b$ ,  $a = b$ ,  $b < a$  can be true.
- *Transitivity*:  $a < b$  and  $b < c \implies a < c$ .
- $(+, <)$  *compatibility*:  $a < b \implies a + c < b + c$ .
- $(\times, <)$  *compatibility*:  $a < b$ ,  $0 < c \implies ac < bc$ .

Because of these properties, the real numbers form an *ordered field*.

In fact, the rational numbers also have the arithmetic operations and the order relation, such that these usual properties are also satisfied. Therefore the rational numbers also form an ordered field. The key distinction between the real and the rational numbers is the existence of *limit*. The issue will be discussed in Section 1.4. Due to this extra property, the real numbers form a *complete ordered field*, while the rational numbers form an incomplete ordered field.

A consequence of the existence of the limit is the existence of the exponential operation  $a^b$  for real numbers  $a > 0$  and  $b$ . In contrast, within the rational numbers, there is no exponential operation, because  $a^b$  may be irrational for rational  $a$  and  $b$ . The exponential of real numbers has many usual properties such as the following.

- *Zero*:  $a^0 = 1$ .
- *Unit*:  $a^1 = a$ ,  $1^a = 1$ .
- *Addition*:  $a^{b+c} = a^b a^c$ .
- *Multiplication*:  $a^{bc} = (a^b)^c$ ,  $(ab)^c = a^c b^c$ .
- *Order*:  $a > b$ ,  $c > 0 \implies a^c > b^c$ ; and  $a > 1$ ,  $b > c \implies a^b > a^c$ .

The exponential operation and the related properties are not assumptions added to the real numbers. They can be derived from the existing arithmetic operations and order relation.

In summary, this course assumes all the knowledge about the arithmetic operations, the exponential operation, and the order relation of real numbers. Starting from Definitions 1.4.1 and 1.4.2 in Section 1.4, we will further assume the existence of limit.

**Exercise 1.1.** Show that the sequence

$$1.4, 1.41, 1.414, 1.4142, 1.41421, 1.414213, 1.4142135, 1.41421356, \dots$$

of more and more refined decimal approximations of  $\sqrt{2}$  converges to  $\sqrt{2}$ . More generally, a positive real number  $a > 0$  has the decimal expansion

$$a = X.Z_1Z_2 \cdots Z_nZ_{n+1} \cdots,$$

where  $X$  is a non-negative integer, and  $Z_n$  is a single digit integer from  $\{0, 1, 2, \dots, 9\}$ . Prove that the sequence

$$X.Z_1, X.Z_1Z_2, X.Z_1Z_2Z_3, X.Z_1Z_2Z_3Z_4, \dots$$

of more and more refined decimal approximations converges to  $a$ .

Exercise 1.2. Suppose  $x_n \leq l \leq y_n$  and  $\lim_{n \rightarrow \infty} (x_n - y_n) = 0$ . Prove that  $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} y_n = l$ .

Exercise 1.3. Suppose  $|x_n - l| \leq y_n$  and  $\lim_{n \rightarrow \infty} y_n = 0$ . Prove that  $\lim_{n \rightarrow \infty} x_n = l$ .

Exercise 1.4. Suppose  $\lim_{n \rightarrow \infty} x_n = l$ . Prove that  $\lim_{n \rightarrow \infty} |x_n| = |l|$ . Is the converse true?

Exercise 1.5. Suppose  $\lim_{n \rightarrow \infty} x_n = l$ . Prove that  $\lim_{n \rightarrow \infty} x_{n+3} = l$ . Is the converse true?

Exercise 1.6. Prove that the limit is not changed if finitely many terms are modified. In other words, if there is  $N$ , such that  $x_n = y_n$  for  $n > N$ , then  $\lim_{n \rightarrow \infty} x_n = l$  if and only if  $\lim_{n \rightarrow \infty} y_n = l$ .

Exercise 1.7. Prove the uniqueness of the limit. In other words, if  $\lim_{n \rightarrow \infty} x_n = l$  and  $\lim_{n \rightarrow \infty} x_n = l'$ , then  $l = l'$ .

Exercise 1.8. Prove the following are equivalent to the definition of  $\lim_{n \rightarrow \infty} x_n = l$ .

1. For any  $c > \epsilon > 0$ , where  $c$  is some fixed number, there is  $N$ , such that  $|x_n - l| < \epsilon$  for all  $n > N$ .
2. For any  $\epsilon > 0$ , there is a natural number  $N$ , such that  $|x_n - l| < \epsilon$  for all  $n > N$ .
3. For any  $\epsilon > 0$ , there is  $N$ , such that  $|x_n - l| \leq \epsilon$  for all  $n > N$ .
4. For any  $\epsilon > 0$ , there is  $N$ , such that  $|x_n - l| < \epsilon$  for all  $n \geq N$ .
5. For any  $\epsilon > 0$ , there is  $N$ , such that  $|x_n - l| \leq 2\epsilon$  for all  $n > N$ .

Exercise 1.9. Which are equivalent to the definition of  $\lim_{n \rightarrow \infty} x_n = l$ ?

1. For  $\epsilon = 0.001$ , we have  $N = 1000$ , such that  $|x_n - l| < \epsilon$  for all  $n > N$ .
2. For any  $0.001 \geq \epsilon > 0$ , there is  $N$ , such that  $|x_n - l| < \epsilon$  for all  $n > N$ .
3. For any  $\epsilon > 0.001$ , there is  $N$ , such that  $|x_n - l| < \epsilon$  for all  $n \geq N$ .
4. For any  $\epsilon > 0$ , there is a natural number  $N$ , such that  $|x_n - l| \leq \epsilon$  for all  $n \geq N$ .
5. For any  $\epsilon > 0$ , there is  $N$ , such that  $|x_n - l| < 2\epsilon^2$  for all  $n > N$ .
6. For any  $\epsilon > 0$ , there is  $N$ , such that  $|x_n - l| < 2\epsilon^2 + 1$  for all  $n > N$ .
7. For any  $\epsilon > 0$ , we have  $N = 1000$ , such that  $|x_n - l| < \epsilon$  for all  $n > N$ .
8. For any  $\epsilon > 0$ , there are infinitely many  $n$ , such that  $|x_n - l| < \epsilon$ .
9. For infinitely many  $\epsilon > 0$ , there is  $N$ , such that  $|x_n - l| < \epsilon$  for all  $n > N$ .
10. For any  $\epsilon > 0$ , there is  $N$ , such that  $l - 2\epsilon < x_n < l + \epsilon$  for all  $n > N$ .
11. For any natural number  $K$ , there is  $N$ , such that  $|x_n - l| < \frac{1}{K}$  for all  $n > N$ .

Exercise 1.10. Write down the complete sets of axioms for the following algebraic structures.

1. Abelian group: A set with addition and subtraction.
2. Field: A set with four arithmetic operations.
3. Ordered field: A set with arithmetic operations and order relation.

## 1.2 Property of Limit

### Boundedness

A sequence is *bounded* if  $|x_n| \leq B$  for some constant  $B$  and all  $n$ . This is equivalent to  $B_1 \leq x_n \leq B_2$  for some constants  $B_1, B_2$  and all  $n$ . The constants  $B, B_1, B_2$  are respectively called a *bound*, a *lower bound* and an *upper bound*.

**Proposition 1.2.1.** *Convergent sequences are bounded.*

*Proof.* Suppose  $x_n$  converges to  $l$ . For  $\epsilon = 1 > 0$ , there is  $N$ , such that

$$n > N \implies |x_n - l| < 1 \iff l - 1 < x_n < l + 1.$$

Moreover, by taking a bigger natural number if necessary, we may further assume  $N$  is a natural number. Then  $x_{N+1}, x_{N+2}, \dots$ , have upper bound  $l + 1$  and lower bound  $l - 1$ , and the whole sequence has upper bound  $\max\{x_1, x_2, \dots, x_N, l + 1\}$  and lower bound  $\min\{x_1, x_2, \dots, x_N, l - 1\}$ .  $\square$

**Example 1.2.1.** The sequences  $n, \sqrt{n}, 2^n, (-3)^{\sqrt{n}}, (1 + (-1)^n)^n$  are not bounded and are therefore divergent.

**Exercise 1.11.** Prove that if  $|x_n| < B$  for  $n > N$ , then the whole sequence  $x_n$  is bounded. This implies that the boundedness is not changed by modifying finitely many terms.

**Exercise 1.12.** Prove that the addition, subtraction and multiplication of bounded sequences are bounded. What about the division and exponential operations? What can you say about the order relation and the boundedness?

**Exercise 1.13.** Suppose  $\lim_{n \rightarrow \infty} x_n = 0$  and  $y_n$  is bounded. Prove that  $\lim_{n \rightarrow \infty} x_n y_n = 0$ .

### Subsequence

A *subsequence* of a sequence  $x_n$  is obtained by selecting some terms. The indices of the selected terms can be arranged as a *strictly increasing* sequence  $n_1 < n_2 < \dots < n_k < \dots$ , and the subsequence can be denoted as  $x_{n_k}$ . The following are two examples of subsequences

$$x_{3k}: x_3, x_6, x_9, x_{12}, x_{15}, x_{18}, \dots,$$

$$x_{2^k}: x_2, x_4, x_8, x_{16}, x_{32}, x_{64}, \dots$$

Note that if  $x_n$  starts from  $n = 1$ , then  $n_k \geq k$ . Therefore by reindexing the terms if necessary, we may always assume  $n_k \geq k$  in subsequent proofs.

**Proposition 1.2.2.** *Suppose a sequence converges to  $l$ . Then all its subsequences converge to  $l$ .*

*Proof.* Suppose  $x_n$  converges to  $l$ . For any  $\epsilon > 0$ , there is  $N$ , such that  $n > N$  implies  $|x_n - l| < \epsilon$ . Then

$$k > N \implies n_k \geq k > N \implies |x_{n_k} - l| < \epsilon. \quad \square$$

**Example 1.2.2.** The sequence  $(-1)^n$  has subsequences  $(-1)^{2k} = 1$  and  $(-1)^{2k+1} = -1$ . Since the two subsequences have different limits, the original sequence diverges. This also gives a counterexample to the converse of Proposition 1.2.1. The right converse of the proposition is given by Theorem 1.5.1.

**Exercise 1.14.** Explain why the sequences diverge.

- |  |  |
|--|--|
| 1. $\sqrt[3]{-n}$ .  | 5. $\frac{n \sin \frac{n\pi}{3}}{n \cos \frac{n\pi}{2} + 2}$ . |
| 2. $\frac{(-1)^n 2n + 1}{n + 2}$ .                                   | 6. $x_{2n} = \frac{1}{n}, x_{2n+1} = \sqrt[n]{n}$ .            |
| 3. $\frac{(-1)^n 2n(n+1)}{(\sqrt{n} + 2)^3}$ .                       | 7. $x_{2n} = 1, x_{2n+1} = \sqrt{n}$ .                         |
| 4. $\sqrt{n} \left( \sqrt{n + (-1)^n} - \sqrt{n - (-1)^n} \right)$ . | 8. $\sqrt[n]{2^n + 3(-1)^n n}$ .                               |

**Exercise 1.15.** Prove that  $\lim_{n \rightarrow \infty} x_n = l$  if and only if  $\lim_{k \rightarrow \infty} x_{2k} = \lim_{k \rightarrow \infty} x_{2k+1} = l$ .

**Exercise 1.16.** Suppose  $x_n$  is the union of two subsequences  $x_{m_k}$  and  $x_{n_k}$ . Prove that  $\lim_{n \rightarrow \infty} x_n = l$  if and only if  $\lim_{k \rightarrow \infty} x_{m_k} = \lim_{k \rightarrow \infty} x_{n_k} = l$ . This extends Exercise 1.15. In general, if a sequence  $x_n$  is the union of finitely many subsequences  $x_{n_{i,k}}$ ,  $i = 1, \dots, p$ ,  $k = 1, 2, \dots$ , then  $\lim_{n \rightarrow \infty} x_n = l$  if and only if  $\lim_{k \rightarrow \infty} x_{n_{i,k}} = l$  for all  $i$ .

## Arithmetic Property

**Proposition 1.2.3.** Suppose  $\lim_{n \rightarrow \infty} x_n = l$  and  $\lim_{n \rightarrow \infty} y_n = k$ . Then

$$\lim_{n \rightarrow \infty} (x_n + y_n) = l + k, \quad \lim_{n \rightarrow \infty} x_n y_n = lk, \quad \lim_{n \rightarrow \infty} \frac{x_n}{y_n} = \frac{l}{k},$$

where  $y_n \neq 0$  and  $k \neq 0$  are assumed in the third equality.

*Proof.* For any  $\epsilon > 0$ , there are  $N_1$  and  $N_2$ , such that

$$\begin{aligned} n > N_1 &\implies |x_n - l| < \frac{\epsilon}{2}, \\ n > N_2 &\implies |y_n - k| < \frac{\epsilon}{2}. \end{aligned}$$

Then for  $n > \max\{N_1, N_2\}$ , we have

$$|(x_n + y_n) - (l + k)| \leq |x_n - l| + |y_n - k| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This completes the proof of  $\lim_{n \rightarrow \infty} (x_n + y_n) = l + k$ .

By Proposition 1.2.1, we have  $|y_n| < B$  for a fixed number  $B$  and all  $n$ . For any  $\epsilon > 0$ , there are  $N_1$  and  $N_2$ , such that

$$\begin{aligned} n > N_1 &\implies |x_n - l| < \frac{\epsilon}{2B}, \\ n > N_2 &\implies |y_n - k| < \frac{\epsilon}{2|l|}. \end{aligned}$$

Then for  $n > \max\{N_1, N_2\}$ , we have

$$\begin{aligned} |x_n y_n - lk| &= |(x_n y_n - l y_n) + (l y_n - lk)| \\ &\leq |x_n - l| |y_n| + |l| |y_n - k| < \frac{\epsilon}{2B} B + |l| \frac{\epsilon}{2|l|} = \epsilon. \end{aligned}$$

This completes the proof of  $\lim_{n \rightarrow \infty} x_n y_n = lk$ .

Assume  $y_n \neq 0$  and  $k \neq 0$ . We will prove  $\lim_{n \rightarrow \infty} \frac{1}{y_n} = \frac{1}{k}$ . Then by the product property of the limit, this implies

$$\lim_{n \rightarrow \infty} \frac{x_n}{y_n} = \lim_{n \rightarrow \infty} x_n \lim_{n \rightarrow \infty} \frac{1}{y_n} = l \frac{1}{k} = \frac{l}{k}.$$

For any  $\epsilon > 0$ , we have  $\epsilon' = \min \left\{ \frac{\epsilon |k|^2}{2}, \frac{|k|}{2} \right\} > 0$ . Then there is  $N$ , such that

$$\begin{aligned} n > N &\implies |y_n - k| < \epsilon' \\ &\iff |y_n - k| < \frac{\epsilon |k|^2}{2}, \quad |y_n - k| < \frac{|k|}{2} \\ &\implies |y_n - k| < \frac{\epsilon |k|^2}{2}, \quad |y_n| > \frac{|k|}{2} \\ &\implies \left| \frac{1}{y_n} - \frac{1}{k} \right| = \frac{|y_n - k|}{|y_n k|} < \frac{\frac{\epsilon |k|^2}{2}}{\frac{|k|}{2} |k|} = \epsilon. \end{aligned}$$

This completes the proof of  $\lim_{n \rightarrow \infty} \frac{1}{y_n} = \frac{1}{k}$ . □

**Exercise 1.17.** Here is another way of proving the limit of quotient.

1. Prove that  $|y - 1| < \epsilon < \frac{1}{2}$  implies  $\left| \frac{1}{y} - 1 \right| < 2\epsilon$ .
2. Prove that  $\lim y_n = 1$  implies  $\lim \frac{1}{y_n} = 1$ .
3. Use the the second part and the limit of multiplication to prove the limit of quotient.

**Exercise 1.18.** Suppose  $\lim_{n \rightarrow \infty} x_n = l$  and  $\lim_{n \rightarrow \infty} y_n = k$ . Prove that

$$\lim_{n \rightarrow \infty} \max\{x_n, y_n\} = \max\{l, k\}, \quad \lim_{n \rightarrow \infty} \min\{x_n, y_n\} = \min\{l, k\}.$$

You may use the formula  $\max\{x, y\} = \frac{1}{2}(x + y + |x - y|)$  and the similar one for  $\min\{x, y\}$ .

**Exercise 1.19.** What is wrong with the following application of Propositions 1.2.2 and 1.2.3: The sequence  $x_n = (-1)^n$  satisfies  $x_{n+1} = -x_n$ . Therefore

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} x_{n+1} = - \lim_{n \rightarrow \infty} x_n,$$

and we get  $\lim_{n \rightarrow \infty} x_n = 0$ .

## Order Property

**Proposition 1.2.4.** *Suppose  $x_n$  and  $y_n$  converge.*

1.  $x_n \geq y_n$  for sufficiently big  $n \implies \lim_{n \rightarrow \infty} x_n \geq \lim_{n \rightarrow \infty} y_n$ .
2.  $\lim_{n \rightarrow \infty} x_n > \lim_{n \rightarrow \infty} y_n \implies x_n > y_n$  for sufficiently big  $n$ .

A special case of the property is that

$$x_n \geq l \text{ for big } n \implies \lim_{n \rightarrow \infty} x_n \geq l,$$

and

$$\lim_{n \rightarrow \infty} x_n > l \implies x_n > l \text{ for big } n.$$

We also have the  $\leq$  and  $<$  versions of the special case.

*Proof.* We prove the second statement first. By Proposition 1.2.3, the assumption implies  $\epsilon = \lim_{n \rightarrow \infty} (x_n - y_n) = \lim_{n \rightarrow \infty} x_n - \lim_{n \rightarrow \infty} y_n > 0$ . Then there is  $N$ , such that

$$n > N \implies |(x_n - y_n) - \epsilon| < \epsilon \implies x_n - y_n - \epsilon > -\epsilon \iff x_n > y_n.$$

By exchanging  $x_n$  and  $y_n$  in the second statement, we find that

$$\lim_{n \rightarrow \infty} x_n < \lim_{n \rightarrow \infty} y_n \implies x_n < y_n \text{ for big } n.$$

This further implies that we cannot have  $x_n \geq y_n$  for big  $n$ . The combined implication

$$\lim_{n \rightarrow \infty} x_n < \lim_{n \rightarrow \infty} y_n \implies \text{opposite of } (x_n \geq y_n \text{ for big } n)$$

is equivalent to the first statement. □

In the second part of the proof above, we used the logical fact that “ $A \implies B$ ” is the same as “(not  $B$ )  $\implies$  (not  $A$ )”. Moreover, we note that the following two statements are not opposite of each other.

- $x_n < y_n$  for big  $n$ : There is  $N$ , such that  $x_n < y_n$  for  $n > N$ .
- $x_n \geq y_n$  for big  $n$ : There is  $N$ , such that  $x_n \geq y_n$  for  $n > N$ .

In fact, the opposite of the second statement is the following: For any  $N$ , there is  $n > N$ , such that  $x_n < y_n$ . The first statement implies (but is not equivalent to) this opposite statement.



## Sandwich Property

**Proposition 1.2.5.** *Suppose*

$$x_n \leq y_n \leq z_n, \quad \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} z_n = l.$$

*Then*  $\lim_{n \rightarrow \infty} y_n = l$ .

*Proof.* For any  $\epsilon > 0$ , there is  $N$ , such that

$$\begin{aligned} n > N &\implies |x_n - l| < \epsilon, |z_n - l| < \epsilon \\ &\implies l - \epsilon < x_n, z_n < l + \epsilon \\ &\implies l - \epsilon < x_n \leq y_n \leq z_n < l + \epsilon \\ &\iff |y_n - l| < \epsilon. \end{aligned}$$

□

In the proof above, we usually have  $N_1$  and  $N_2$  for  $\lim_{n \rightarrow \infty} x_n$  and  $\lim_{n \rightarrow \infty} z_n$  respectively. Then  $N = \max\{N_1, N_2\}$  works for both limits. In the later arguments, we may always choose the same  $N$  for finitely many limits.

**Example 1.2.3.** For any  $a > 1$  and  $n > a$ , we have  $1 < \sqrt[n]{a} < \sqrt[n]{n}$ . Then by the limit (1.1.3) and the sandwich rule, we have  $\lim_{n \rightarrow \infty} \sqrt[n]{a} = 1$ . On the other hand, for  $0 < a < 1$ , we have  $b = \frac{1}{a} > 1$  and

$$\lim_{n \rightarrow \infty} \sqrt[n]{a} = \lim_{n \rightarrow \infty} \frac{1}{\sqrt[n]{b}} = \frac{1}{\lim_{n \rightarrow \infty} \sqrt[n]{b}} = 1.$$

Combining all the cases, we get  $\lim_{n \rightarrow \infty} \sqrt[n]{a} = 1$  for any  $a > 0$ .

**Exercise 1.20.** Redo Exercise 1.3 by using the sandwich rule.

**Exercise 1.21.** Let  $a > 0$  be a constant. Then  $\frac{1}{n} < a < n$  for big  $n$ . Use this and the limit (1.1.3) to prove  $\lim_{n \rightarrow \infty} \sqrt[n]{a} = 1$ .

## 1.3 Infinity and Infinitesimal

A changing numerical quantity is an *infinity* if it tends to get arbitrarily big. For sequences, this means the following.

**Definition 1.3.1.** A sequence  $x_n$  diverges to *infinity*, denoted  $\lim_{n \rightarrow \infty} x_n = \infty$ , if for any  $b$ , there is  $N$ , such that

$$n > N \implies |x_n| > b. \quad (1.3.1)$$

It diverges to *positive infinity*, denoted  $\lim_{n \rightarrow \infty} x_n = +\infty$ , if for any  $b$ , there is  $N$ , such that

$$n > N \implies x_n > b.$$

It diverges to *negative infinity*, denoted  $\lim_{n \rightarrow \infty} x_n = -\infty$ , if for any  $b$ , there is  $N$ , such that

$$n > N \implies x_n < b.$$

A changing numerical quantity is an *infinitesimal* if it tends to get arbitrarily small. For sequences, this means that for any  $\epsilon > 0$ , there is  $N$ , such that

$$n > N \implies |x_n| < \epsilon. \quad (1.3.2)$$

This is the same as  $\lim_{n \rightarrow \infty} x_n = 0$ .

Note that the implications (1.3.1) and (1.3.2) are equivalent by changing  $x_n$  to  $\frac{1}{x_n}$  and taking  $\epsilon = \frac{1}{b}$ . Therefore we have

$$x_n \text{ is an infinity} \iff \frac{1}{x_n} \text{ is an infinitesimal.}$$

We also note that, since  $\lim_{n \rightarrow \infty} x_n = l$  is equivalent to  $\lim_{n \rightarrow \infty} (x_n - l) = 0$ , we have

$$x_n \text{ converges to } l \iff x_n - l \text{ is an infinitesimal.}$$

**Exercise 1.22.** Infinities must be unbounded. Is the converse true?

**Exercise 1.23.** Prove that if a sequence diverges to infinity, then all its subsequences diverge to infinity.

**Exercise 1.24.** Suppose  $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} y_n = +\infty$ . Prove that  $\lim_{n \rightarrow \infty} (x_n + y_n) = +\infty$  and  $\lim_{n \rightarrow \infty} x_n y_n = +\infty$ .

**Exercise 1.25.** Suppose  $\lim_{n \rightarrow \infty} x_n = \infty$  and  $|x_n - x_{n+1}| < c$  for some constant  $c$ .

1. Prove that either  $\lim_{n \rightarrow \infty} x_n = +\infty$  or  $\lim_{n \rightarrow \infty} x_n = -\infty$ .
2. If we further know  $\lim_{n \rightarrow \infty} x_n = +\infty$ , prove that for any  $a > x_1$ , some term  $x_n$  lies in the interval  $(a, a + c)$ .

**Exercise 1.26.** Prove that if  $\lim_{n \rightarrow \infty} x_n = +\infty$  and  $|x_n - x_{n+1}| < c$  for some constant  $c < \pi$ , then  $\sin x_n$  diverges. Exercise 1.25 might be helpful here.

Some properties of finite limits can be extended to infinities and infinitesimals. For example, the properties in Exercise 1.24 can be denoted as the arithmetic rules  $(+\infty) + (+\infty) = +\infty$  and  $(+\infty)(+\infty) = +\infty$ . Moreover, if  $\lim_{n \rightarrow \infty} x_n = 1$ ,  $\lim_{n \rightarrow \infty} y_n = 0$ , and  $y_n < 0$  for big  $n$ , then  $\lim_{n \rightarrow \infty} \frac{x_n}{y_n} = -\infty$ . Thus we have another arithmetic rule  $\frac{1}{0^-} = -\infty$ . Common sense suggests more arithmetic rules such as

$$\begin{array}{lll} c + \infty = \infty, & c \cdot \infty = \infty \text{ for } c \neq 0, & \infty \cdot \infty = \infty, \\ \frac{\infty}{c} = \infty, & \frac{c}{0} = \infty \text{ for } c \neq 0, & \frac{c}{\infty} = 0, \end{array}$$

where  $c$  is a finite number and represents a sequence converging to  $c$ . On the other hand, we must be careful not to overextend the arithmetic rules. The following example shows that  $\frac{0}{0}$  has no definite value.

$$\begin{array}{lll} \lim_{n \rightarrow \infty} n^{-1} = 0, & \lim_{n \rightarrow \infty} 2n^{-1} = 0, & \lim_{n \rightarrow \infty} n^{-2} = 0, \\ \lim_{n \rightarrow \infty} \frac{n^{-1}}{2n^{-1}} = \frac{1}{2}, & \lim_{n \rightarrow \infty} \frac{n^{-1}}{n^{-2}} = +\infty, & \lim_{n \rightarrow \infty} \frac{n^{-2}}{2n^{-1}} = 0. \end{array}$$

**Exercise 1.27.** Prove properties of infinity.

1. (bounded) $+\infty = \infty$ : If  $x_n$  is bounded and  $\lim_{n \rightarrow \infty} y_n = \infty$ , then  $\lim_{n \rightarrow \infty} (x_n + y_n) = \infty$ .
2.  $(-\infty)(-\infty) = +\infty$ .
3.  $\min\{+\infty, +\infty\} = +\infty$ .
4. Sandwich rule: If  $x_n \geq y_n$  and  $\lim_{n \rightarrow \infty} y_n = +\infty$ , then  $\lim_{n \rightarrow \infty} x_n = +\infty$ .
5.  $(> c > 0) \cdot (+\infty) = +\infty$ : If  $x_n > c$  for some constant  $c > 0$  and  $\lim_{n \rightarrow \infty} y_n = +\infty$ , then  $\lim_{n \rightarrow \infty} x_n y_n = +\infty$ .

**Exercise 1.28.** Show that  $\infty + \infty$  has no definite value by constructing examples of sequences  $x_n$  and  $y_n$  that diverge to  $\infty$  but one of the following holds.

1.  $\lim_{n \rightarrow \infty} (x_n + y_n) = 2$ .
2.  $\lim_{n \rightarrow \infty} (x_n + y_n) = +\infty$ .
3.  $x_n + y_n$  is bounded and divergent.

**Exercise 1.29.** Show that  $0 \cdot \infty$  has no definite value by constructing examples of sequences  $x_n$  and  $y_n$ , such that  $\lim_{n \rightarrow \infty} x_n = 0$  and  $\lim_{n \rightarrow \infty} y_n = \infty$  and one of the following holds.

1.  $\lim_{n \rightarrow \infty} x_n y_n = 2$ .
2.  $\lim_{n \rightarrow \infty} x_n y_n = 0$ .
3.  $\lim_{n \rightarrow \infty} x_n y_n = \infty$ .
4.  $x_n y_n$  is bounded and divergent.

**Exercise 1.30.** Provide counterexamples to the wrong arithmetic rules.

$$\frac{+\infty}{+\infty} = 1, \quad (+\infty) - (+\infty) = 0, \quad 0 \cdot \infty = 0, \quad 0 \cdot \infty = \infty, \quad 0 \cdot \infty = 1.$$

## 1.4 Supremum and Infimum

Both real numbers  $\mathbb{R}$  and rational numbers  $\mathbb{Q}$  are ordered fields. The deeper part of the mathematical analysis lies in the difference between the real and rational numbers. The key difference is the existence of limit, which is the same as the completeness of order.

**Definition 1.4.1.** Let  $X$  be a nonempty set of numbers. An *upper bound* of  $X$  is a number  $B$  such that  $x \leq B$  for any  $x \in X$ . The *supremum* of  $X$  is the least upper bound of the set and is denoted  $\sup X$ .

**Definition 1.4.2.** Real numbers  $\mathbb{R}$  is a set with the usual arithmetic operations and the order satisfying the usual properties, and the additional property that any bounded set of real numbers has the supremum.

In contrast, rational numbers  $\mathbb{Q}$  does not have the additional property. For example, the decimal approximations of  $\sqrt{2}$  in Exercise 1.1 is a bounded rational sequence. The sequence has no rational supremum, because its supremum  $\sqrt{2}$  is not a rational number.

## Supremum and Infimum

The supremum  $\lambda = \sup X$  is characterized by the following properties.

1.  $\lambda$  is an upper bound: For any  $x \in X$ , we have  $x \leq \lambda$ .
2. Any number smaller than  $\lambda$  is not an upper bound: For any  $\epsilon > 0$ , there is  $x \in X$ , such that  $x > \lambda - \epsilon$ .

The *infimum*  $\inf X$  is the greatest lower bound, and can be similarly characterized.

**Example 1.4.1.** Both the set  $\{1, 2\}$  and the interval  $[0, 2]$  have 2 as the supremum. In general, the *maximum* of a set  $X$  is a number  $\xi \in X$  satisfying  $\xi \geq x$  for any  $x \in X$ . If the maximum exists, then the maximum is the supremum. We also note that the interval  $(0, 2)$  has no maximum but still has 2 as the supremum.

A bounded set of real numbers may not always have maximum, but always has supremum.

Similarly, the *minimum* of a set  $X$  is a number  $\eta \in X$  satisfying  $\eta \leq x$  for any  $x \in X$ . If the minimum exists, then it is the infimum.

**Example 1.4.2.** The irrational number  $\sqrt{2}$  is the supremum of the set

$$\{1.4, 1.41, 1.414, 1.4142, 1.41421, 1.414213, 1.4142135, 1.41421356, \dots\}$$

of its decimal expansions. It is also the supremum of the set

$$\left\{ \frac{m}{n} : m \text{ and } n \text{ are natural numbers satisfying } m^2 < 2n^2 \right\}$$

of positive rational numbers whose squares are less than 2.

**Example 1.4.3.** Let  $L_n$  be the length of an edge of the inscribed regular  $n$ -gon in a circle of radius 1. Then  $2\pi$  is the supremum of the set  $\{3L_3, 4L_4, 5L_5, \dots\}$  of the circumferences of the inscribed regular  $n$ -gons.

**Exercise 1.31.** Find the suprema and the infima.

1.  $\{a + b : a, b \text{ are rational, } a^2 < 3, |2b + 1| < 5\}$ .
2.  $\left\{ \frac{n}{n+1} : n \text{ is a natural number} \right\}$ .
3.  $\left\{ \frac{(-1)^n n}{n+1} : n \text{ is a natural number} \right\}$ .

4.  $\left\{\frac{m}{n} : m \text{ and } n \text{ are natural numbers satisfying } m^2 > 3n^2\right\}$ .
5.  $\left\{\frac{1}{2^m} + \frac{1}{3^n} : m \text{ and } n \text{ are natural numbers}\right\}$ .
6.  $\{nR_n : n \geq 3 \text{ is a natural number}\}$ , where  $R_n$  is the length of an edge of the circumscribed regular  $n$ -gon around a circle of radius 1.

**Exercise 1.32.** Prove that the supremum is unique.

**Exercise 1.33.** Suppose  $X$  is a nonempty bounded set of numbers. Prove that  $\lambda = \sup X$  is characterized by the following two properties.

1.  $\lambda$  is an upper bound: For any  $x \in X$ , we have  $x \leq \lambda$ .
2.  $\lambda$  is the limit of a sequence in  $X$ : There are  $x_n \in X$ , such that  $\lambda = \lim_{n \rightarrow \infty} x_n$ .

The following are some properties of the supremum and infimum.

**Proposition 1.4.3.** *Suppose  $X$  and  $Y$  are nonempty bounded sets of numbers.*

1.  $\sup X \geq \inf X$ , and the equality holds if and only if  $X$  contains a single number.
2.  $\sup X \leq \inf Y$  if and only if  $x \leq y$  for any  $x \in X$  and  $y \in Y$ .
3.  $\sup X \geq \inf Y$  if and only if for any  $\epsilon > 0$ , there are  $x \in X$  and  $y \in Y$  satisfying  $y - x < \epsilon$ .
4. Let  $X + Y = \{x + y : x \in X, y \in Y\}$ . Then  $\sup(X + Y) = \sup X + \sup Y$  and  $\inf(X + Y) = \inf X + \inf Y$ .
5. Let  $cX = \{cx : x \in X\}$ . Then  $\sup(cX) = c\sup X$  when  $c > 0$  and  $\sup(cX) = c\inf X$  when  $c < 0$ . In particular,  $\sup(-X) = -\inf X$ .
6. Let  $XY = \{xy : x \in X, y \in Y\}$ . If all numbers in  $X, Y$  are positive, then  $\sup(XY) = \sup X \sup Y$  and  $\inf(XY) = \inf X \inf Y$ .
7. Let  $X^{-1} = \{x^{-1} : x \in X\}$ . If all numbers in  $X$  are positive, then  $\sup X^{-1} = (\inf X)^{-1}$ .
8.  $|x - y| \leq c$  for any  $x \in X$  and  $y \in Y$  if and only if  $|\sup X - \sup Y| \leq c$ ,  $|\inf X - \inf Y| \leq c$ ,  $|\sup X - \inf Y| \leq c$  and  $|\inf X - \sup Y| \leq c$ .

*Proof.* For the first property, we pick any  $x \in X$  and get  $\sup x \geq x \geq \inf X$ . When  $\sup x = \inf X$ , we get  $x = \sup x = \inf X$ , so that  $X$  contains a single number.

For the second property,  $\sup X \leq \inf Y$  implies  $x \leq \sup X \leq \inf Y \leq y$  for any  $x \in X$  and  $y \in Y$ . Conversely, assume  $x \leq y$  for any  $x \in X$  and  $y \in Y$ . Then any  $y \in Y$  is an upper bound of  $X$ . Therefore  $\sup X \leq y$ . Since  $\sup X \leq y$  for any  $y \in Y$ ,  $\sup X$  is a lower bound of  $Y$ . Therefore  $\sup X \leq \inf Y$ .

For the third property, for any  $\epsilon > 0$ , there are  $x \in X$  satisfying  $x > \sup X - \frac{\epsilon}{2}$  and  $y \in Y$  satisfying  $y < \inf Y + \frac{\epsilon}{2}$ . If  $\sup X \geq \inf Y$ , then this implies  $y - x <$

$(\inf Y + \frac{\epsilon}{2}) - (\sup X - \frac{\epsilon}{2}) = \inf Y - \sup X + \epsilon \leq \epsilon$ . Conversely, suppose for any  $\epsilon > 0$ , there are  $x \in X$  and  $y \in Y$  satisfying  $y - x < \epsilon$ . Then  $\sup X - \inf Y \geq x - y \geq -\epsilon$ . Since  $\epsilon > 0$  is arbitrary, we get  $\sup X - \inf Y \geq 0$ .

For the fourth property, we have  $x + y \leq \sup X + \sup Y$  for any  $x \in X$  and  $y \in Y$ . This means  $\sup X + \sup Y$  is an upper bound of  $X + Y$ . Moreover, for any  $\epsilon > 0$ , there are  $x \in X$  and  $y \in Y$  satisfying  $x > \sup X - \frac{\epsilon}{2}$  and  $y > \sup Y - \frac{\epsilon}{2}$ . Then  $x + y \in X + Y$  satisfies  $x + y > \sup X + \sup Y - \epsilon$ . Thus the two conditions for  $\sup X + \sup Y$  to be the supremum of  $X + Y$  are verified.

For the fifth property, assume  $c > 0$ . Then  $l > x$  for all  $x \in X$  is the same as  $cl > cx$  for all  $x \in X$ . Therefore  $l$  is an upper bound of  $X$  if and only if  $cl$  is an upper bound of  $cX$ . In particular,  $l$  is the smallest upper bound of  $X$  if and only if  $cl$  is the smallest upper bound of  $cX$ . This means  $\sup(cX) = c \sup X$ .

On the other hand, assume  $c < 0$ . Then  $l < x$  for all  $x \in X$  is the same as  $cl > cx$  for all  $x \in X$ . Therefore  $l$  is a lower bound of  $X$  if and only if  $cl$  is an upper bound of  $cX$ . This means  $\sup(cX) = c \inf X$ .

The proof of the last three properties are left to the reader.  $\square$

**Exercise 1.34.** Finish the proof of Proposition 1.4.3.

**Exercise 1.35.** Suppose  $X_i$  are nonempty sets of numbers. Let  $X = \cup_i X_i$  and  $\lambda_i = \sup X_i$ . Prove that  $\sup X = \sup_i \lambda_i = \sup_i \sup X_i$ . What about the infimum?

## Monotone Sequence

As the time goes by, the world record in 100 meter dash is shorter and shorter time, and the limit should be the infimum of all the world records. The example suggests that a bounded (for the existence of infimum) decreasing sequence should converge to its infimum.

A sequence  $x_n$  is *increasing* if  $x_{n+1} \geq x_n$  (for all  $n$ ). It is *strictly increasing* if  $x_{n+1} > x_n$ . The concepts of *decreasing* and *strictly decreasing* sequences can be similarly defined. A sequence is *monotone* if it is either increasing or decreasing.

**Proposition 1.4.4.** *A bounded monotone sequence of real numbers converges. An unbounded monotone sequence of real numbers diverges to infinity.*

*Proof.* A bounded and increasing sequence  $x_n$  has a real number supremum  $l = \sup\{x_n\}$ . For any  $\epsilon > 0$ , by the second property that characterizes the supremum, there is  $N$ , such that  $x_N > l - \epsilon$ . Since the sequence is increasing,  $n > N$  implies  $x_n \geq x_N > l - \epsilon$ . We also have  $x_n \leq l$  because  $l$  is an upper bound. Therefore we conclude that

$$n > N \implies l - \epsilon < x_n \leq l \implies |x_n - l| < \epsilon.$$

This proves that the sequence converges to  $l$ .

If  $x_n$  is unbounded and increasing, then it has no upper bound (see Exercise 1.36). In other words, for any  $b$ , there is  $N$ , such that  $x_N > b$ . Since the sequence

is increasing, we have

$$n > N \implies x_n \geq x_N > b.$$

This proves that the sequence diverges to  $+\infty$ .

The proof for a decreasing sequence is similar.  $\square$

**Example 1.4.4.** The natural constant  $e$  is defined as the limit

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = 2.71828182845904 \dots \quad (1.4.1)$$

Here we justify the definition by showing that the limit converges.

Let  $x_n = \left(1 + \frac{1}{n}\right)^n$ . The binomial expansion tells us

$$\begin{aligned} x_n &= 1 + n \left(\frac{1}{n}\right) + \frac{n(n-1)}{2!} \left(\frac{1}{n}\right)^2 + \frac{n(n-1)(n-2)}{3!} \left(\frac{1}{n}\right)^3 \\ &\quad + \dots + \frac{n(n-1) \cdots 1}{n!} \left(\frac{1}{n}\right)^n \\ &= 1 + \frac{1}{1!} + \frac{1}{2!} \left(1 - \frac{1}{n}\right) + \frac{1}{3!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \\ &\quad + \dots + \frac{1}{n!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{n-1}{n}\right). \end{aligned}$$

By comparing the similar formula for  $x_{n+1}$ , we find the sequence is strictly increasing. The formula also tells us

$$\begin{aligned} x_n &< 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!} \\ &< 1 + \frac{1}{1!} + \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \dots + \frac{1}{(n-1)n} \\ &= 2 + \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \dots + \left(\frac{1}{n-1} - \frac{1}{n}\right) \\ &= 2 + \frac{1}{1} - \frac{1}{n} < 3. \end{aligned}$$

Therefore the sequence converges.

**Exercise 1.36.** Prove that an increasing sequence is bounded if and only if it has an upper bound.

**Exercise 1.37.** Prove that  $\lim_{n \rightarrow \infty} a^n = 0$  for  $|a| < 1$  in the following steps.

1. Prove that for  $0 \leq a < 1$ , the sequence  $a^n$  decreases.
2. Prove that for  $0 \leq a < 1$ , the limit of  $a^n$  must be 0.
3. Prove that  $\lim_{n \rightarrow \infty} a^n = 0$  for  $-1 < a \leq 0$ .

## 1.5 Convergent Subsequence

### Bolzano-Weierstrass Theorem

Proposition 1.2.1 says any convergent sequence is bounded. The counterexample  $(-1)^n$  shows that the converse of the proposition is not true. Despite the divergence, we still note that the sequence is made up of two convergent subsequences  $(-1)^{2k-1} = -1$  and  $(-1)^{2k} = 1$ .

**Theorem 1.5.1** (Bolzano<sup>1</sup>-Weierstrass<sup>2</sup> Theorem). *A bounded sequence of real numbers has a convergent subsequence.*

By Proposition 1.2.2, any subsequence of a convergent sequence is convergent. The theorem says that, if the original sequence is only assumed to be bounded, then “any subsequence” should be changed to “some subsequence”.

*Proof.* The bounded sequence  $x_n$  lies in a bounded interval  $[a, b]$ . Divide  $[a, b]$  into two equal halves  $\left[a, \frac{a+b}{2}\right]$  and  $\left[\frac{a+b}{2}, b\right]$ . Then one of the halves must contain infinitely many  $x_n$ . We denote this interval by  $[a_1, b_1]$ .

Further divide  $[a_1, b_1]$  into two equal halves  $\left[a_1, \frac{a_1+b_1}{2}\right]$  and  $\left[\frac{a_1+b_1}{2}, b_1\right]$ . Again one of the halves, which we denote by  $[a_2, b_2]$ , contains infinitely many  $x_n$ . Keep going, we get a sequence of intervals

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \cdots \supset [a_k, b_k] \supset \cdots$$

with the length  $b_k - a_k = \frac{b-a}{2^k}$ . Moreover, each interval  $[a_k, b_k]$  contains infinitely many  $x_n$ .

The inclusion relation between the intervals implies

$$a \leq a_1 \leq a_2 \leq \cdots \leq a_k \leq \cdots \leq b_k \leq \cdots \leq b_2 \leq b_1 \leq b.$$

Since  $a_k$  and  $b_k$  are also bounded, by Proposition 1.4.4, both sequences converge. Moreover, by Example 1.1.3, we have

$$\lim_{k \rightarrow \infty} (b_k - a_k) = (b - a) \lim_{k \rightarrow \infty} \frac{1}{2^k} = 0.$$

<sup>1</sup>Bernard Placidus Johann Nepomuk Bolzano, born October 5, 1781, died December 18, 1848 in Prague, Bohemia (now Czech). Bolzano is famous for his 1837 book “Theory of Science”. He insisted that many results which were thought “obvious” required rigorous proof and made fundamental contributions to the foundation of mathematics. He understood the need to redefine and enrich the concept of number itself and define the Cauchy sequence four years before Cauchy’s work appeared.

<sup>2</sup>Karl Theodor Wilhelm Weierstrass, born October 31, 1815 in Ostenfelde, Westphalia (now Germany), died February 19, 1848 in Berlin, Germany. In 1864, he found a continuous but nowhere differentiable function. His lectures on analytic functions, elliptic functions, abelian functions and calculus of variations influenced many generations of mathematicians, and his approach still dominates the teaching of analysis today.



Therefore  $a_k$  and  $b_k$  converge to the same limit  $l$ .

Since  $[a_1, b_1]$  contains infinitely many  $x_n$ , we have  $x_{n_1} \in [a_1, b_1]$  for some  $n_1$ . Since  $[a_2, b_2]$  contains infinitely many  $x_n$ , we also have  $x_{n_2} \in [a_2, b_2]$  for some  $n_2 > n_1$ . Keep going, we have a subsequence satisfying  $x_{n_k} \in [a_k, b_k]$ . This means that  $a_k \leq x_{n_k} \leq b_k$ . By the sandwich rule, we get  $\lim_{k \rightarrow \infty} x_{n_k} = l$ . Thus  $x_{n_k}$  is a converging subsequence.  $\square$

The following useful remark is used in the proof above. Suppose  $P$  is a property about terms in a sequence (say  $x_n > l$ , or  $x_n > x_{n+1}$ , or  $x_n \in [a, b]$ , for examples). Then the following statements are equivalent.

1. There are infinitely many  $x_n$  with property  $P$ .
2. For any  $N$ , there is  $n > N$ , such that  $x_n$  has property  $P$ .
3. There is a subsequence  $x_{n_k}$ , such that each term  $x_{n_k}$  has property  $P$ .

**Exercise 1.38.** Explain that any real number is the limit of a sequence of the form  $\frac{n_1}{10}, \frac{n_2}{100}, \frac{n_3}{1000}, \dots$ , where  $n_k$  are integers. Based on this observation, construct a sequence such that any real number in  $[0, 1]$  is the limit of a convergent subsequence.

**Exercise 1.39.** Prove that a number is the limit of a convergent subsequence of  $x_n$  if and only if it is the limit of a convergent subsequence of  $\sqrt[n]{n}x_n$ .

**Exercise 1.40.** Suppose  $x_n$  and  $y_n$  are two bounded sequences. Prove that there are  $n_k$ , such that both subsequences  $x_{n_k}$  and  $y_{n_k}$  converge. Moreover, extend this to more than two bounded sequences.

## Cauchy Criterion

The definition of convergence involves the explicit value of the limit. However, there are many cases that a sequence must be convergent, but the limit value is not known. The limit of the world record in 100 meter dash is one such example. In such cases, the convergence cannot be established by using the definition alone.

**Theorem 1.5.2 (Cauchy<sup>3</sup> Criterion).** *A sequence  $x_n$  converges if and only if for any  $\epsilon > 0$ , there is  $N$ , such that*

$$m, n > N \implies |x_m - x_n| < \epsilon.$$

A sequence  $x_n$  satisfying the condition in the theorem is called *Cauchy sequence*. The theorem says that a sequence converges if and only if it is a Cauchy sequence.

---

<sup>3</sup>Augustin Louis Cauchy, born 1789 in Paris (France), died 1857 in Sceaux (France). His contributions to mathematics can be seen by the numerous mathematical terms bearing his name, including Cauchy integral theorem (complex functions), Cauchy-Kovalevskaya theorem (differential equations), Cauchy-Riemann equations, Cauchy sequences. He produced 789 mathematics papers and his collected works were published in 27 volumes.

*Proof.* Suppose  $x_n$  converges to  $l$ . For any  $\epsilon > 0$ , there is  $N$ , such that  $n > N$  implies  $|x_n - l| < \frac{\epsilon}{2}$ . Then  $m, n > N$  implies

$$|x_m - x_n| = |(x_m - l) - (x_n - l)| \leq |x_m - l| + |x_n - l| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Conversely, for a Cauchy sequence, we prove the convergence in three steps.

1. A Cauchy sequence is bounded.
2. Bounded sequence has converging subsequence.
3. If a Cauchy sequence has a convergent subsequence, then the whole Cauchy sequence converges.

Suppose  $x_n$  is a Cauchy sequence. For  $\epsilon = 1 > 0$ , there is  $N$ , such that  $m, n > N$  implies  $|x_m - x_n| < 1$ . Taking  $m = N + 1$ , we find  $n > N$  implies  $x_{N+1} - 1 < x_n < x_{N+1} + 1$ . Therefore  $\max\{x_1, x_2, \dots, x_N, x_{N+1} + 1\}$  and  $\min\{x_1, x_2, \dots, x_N, x_{N+1} - 1\}$  are upper and lower bounds for the sequence.

By Bolzano-Weierstrass Theorem, there is a subsequence  $x_{n_k}$  converging to a limit  $l$ . This means that for any  $\epsilon > 0$ , there is  $K$ , such that

$$k > K \implies |x_{n_k} - l| < \frac{\epsilon}{2}.$$

On the other hand, since  $x_n$  is a Cauchy sequence, there is  $N$ , such that

$$m, n > N \implies |x_m - x_n| < \frac{\epsilon}{2}.$$

Now for any  $n > N$ , we can easily find some  $k > K$ , such that  $n_k > N$  ( $k = \max\{K, N\} + 1$ , for example). Then we have both  $|x_{n_k} - l| < \frac{\epsilon}{2}$  and  $|x_{n_k} - x_n| < \frac{\epsilon}{2}$ . The inequalities imply  $|x_n - l| < \epsilon$ , and we established the implication

$$n > N \implies |x_n - l| < \epsilon. \quad \square$$

We note that the first and third steps are general facts in any metric space, and the Bolzano-Weierstrass Theorem is only used in the second step. Therefore the equivalence between Cauchy property and convergence remains true in any general setting as long as Bolzano-Weierstrass Theorem remains true.

**Example 1.5.1.** Consider the sequence  $x_n = (-1)^n$ . For  $\epsilon = 1 > 0$  and any  $N$ , we can find an even  $n > N$ . Then  $m = n + 1 > N$  is odd and  $|x_m - x_n| = 2 > \epsilon$ . Therefore the Cauchy criterion fails and the sequence diverges.

**Example 1.5.2 (Oresme<sup>4</sup>).** The *harmonic sequence*

$$x_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}$$

---

<sup>4</sup>Nicole Oresme, born 1323 in Allemagne (France), died 1382 in Lisieux (France). Oresme is best known as an economist, mathematician, and a physicist. He was one of the most famous and influential philosophers of the later Middle Ages. His contributions to mathematics were mainly contained in his manuscript *Tractatus de configuratione qualitatum et motuum* (Treatise on the Configuration of Qualities and Motions).

satisfies

$$x_{2n} - x_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} \geq \frac{1}{2n} + \frac{1}{2n} + \cdots + \frac{1}{2n} = \frac{n}{2n} = \frac{1}{2}.$$

Thus for  $\epsilon = \frac{1}{2}$  and any  $N$ , we have  $|x_m - x_n| > \frac{1}{2}$  by taking any natural number  $n > N$  and  $m = 2n$ . Therefore the Cauchy criterion fails and the harmonic sequence diverges.

**Example 1.5.3.** We show that  $x_n = \sin n$  diverges. For any integer  $k$ , the intervals  $\left(2k\pi + \frac{\pi}{4}, 2k\pi + \frac{3\pi}{4}\right)$  and  $\left(2k\pi - \frac{\pi}{4}, 2k\pi - \frac{3\pi}{4}\right)$  have length  $\frac{\pi}{2} > 1$  and therefore must contain integers  $m_k$  and  $n_k$ . Moreover, by taking  $k$  to be a big positive number,  $m_k$  and  $n_k$  can be as big as we wish. Then  $\sin m_k > \frac{1}{\sqrt{2}}$ ,  $\sin n_k < -\frac{1}{\sqrt{2}}$ , and we have  $|\sin m_k - \sin n_k| > \sqrt{2}$ . Thus the sequence  $\sin n$  is not Cauchy and must diverge.

For extensions of the example, see Exercises 1.42 and 1.26.

**Exercise 1.41.** For the harmonic sequence  $x_n$ , use  $x_{2n} - x_n > \frac{1}{2}$  to prove that  $x_{2n} > \frac{n}{2}$ . Then show the divergence of  $x_n$ .

**Exercise 1.42.** For  $0 < a < \pi$ , prove that both  $\sin na$  and  $\cos na$  diverge.

**Exercise 1.43.** Prove any Cauchy sequence is bounded.

**Exercise 1.44.** Prove that a subsequence of a Cauchy sequence is still a Cauchy sequence.

## Set of Limits

By Bolzano-Weierstrass Theorem, for a bounded sequence  $x_n$ , the set  $\text{LIM}\{x_n\}$  of all the limits of convergent subsequences is not empty. The numbers in  $\text{LIM}\{x_n\}$  are characterized below.

**Proposition 1.5.3.** *A real number  $l$  is the limit of a convergent subsequence of  $x_n$  if and only if for any  $\epsilon > 0$ , there are infinitely many  $x_n$  satisfying  $|x_n - l| < \epsilon$ .*

We remark that the criterion means that, for any  $\epsilon > 0$  and  $N$ , there is  $n > N$ , such that  $|x_n - l| < \epsilon$ .

*Proof.* Suppose  $l$  is the limit of a subsequence  $x_{n_k}$ . For any  $\epsilon > 0$ , there is  $K$ , such that  $k > K$  implies  $|x_{n_k} - l| < \epsilon$ . Then  $x_{n_k}$  for all  $k > K$  are the infinitely many  $x_n$  satisfying  $|x_n - l| < \epsilon$ .

Conversely, suppose for any  $\epsilon > 0$ , there are infinitely many  $x_n$  satisfying  $|x_n - l| < \epsilon$ . Then for  $\epsilon = 1$ , there is  $n_1$  satisfying  $|x_{n_1} - l| < 1$ . Next, for  $\epsilon = \frac{1}{2}$ , there are infinitely many  $x_n$  satisfying  $|x_n - l| < \frac{1}{2}$ . Among these we can find  $n_2 > n_1$ , such that  $|x_{n_2} - l| < \frac{1}{2}$ . After finding  $x_{n_k}$ , we have infinitely many  $x_n$  satisfying  $|x_n - l| < \frac{1}{k+1}$ . Among these we can find  $n_{k+1} > n_k$ , such

that  $|x_{n_{k+1}} - l| < \frac{1}{k+1}$ . We inductively constructed a subsequence  $x_{n_k}$  satisfying  $|x_{n_k} - l| < \frac{1}{k}$ . The inequality implies that  $x_{n_k}$  converges to  $l$ .  $\square$

**Example 1.5.4.** Let  $x_n, y_n, z_n$  be sequences converging to  $l_1, l_2, l_3$ , respectively. Then the set LIM of limits of the sequence

$$x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3, \dots, x_n, y_n, z_n, \dots$$

contains  $l_1, l_2, l_3$ . We claim that  $\text{LIM} = \{l_1, l_2, l_3\}$ .

We need to explain that any  $l \neq l_1, l_2, l_3$  is not the limit of any subsequence. Pick an  $\epsilon > 0$  satisfying

$$|l - l_1| \geq 2\epsilon, \quad |l - l_2| \geq 2\epsilon, \quad |l - l_3| \geq 2\epsilon.$$

Then there is  $N$ , such that

$$n > N \implies |x_n - l_1| < \epsilon, \quad |y_n - l_2| < \epsilon, \quad |z_n - l_3| < \epsilon.$$

Since  $|l - l_1| \geq 2\epsilon$  and  $|x_n - l_1| < \epsilon$  imply  $|x_n - l| > \epsilon$ , we have

$$n > N \implies |x_n - l| > \epsilon, \quad |y_n - l| > \epsilon, \quad |z_n - l| > \epsilon.$$

This implies that  $l$  cannot be the limit of any convergent subsequence.

A more direct argument is the following. Let  $l$  be the limit of a convergent subsequence  $w_m$  of the combined sequence. The subsequence  $w_m$  must contain infinitely many terms from at least one of the three sequences. If  $w_m$  contains infinitely many terms from  $x_n$ , then it contains a subsequence  $x_{n_k}$ , and we get

$$l = \lim w_m = \lim x_{n_k} = \lim x_n = l_1.$$

Here Proposition 1.2.2 is applied to the second and the third equalities.

**Example 1.5.5.** In Example 1.5.3, we already know that  $\sin n$  diverges. Here we show that 0 is the limit of some converging subsequence.

The Hurwitz Theorem in number theory says that, for any irrational number such as  $\pi$ , there are infinitely many rational numbers  $\frac{n}{m}$ ,  $m, n \in \mathbb{Z}$ , such that  $\left| \frac{n}{m} - \pi \right| \leq \frac{1}{\sqrt{5}m^2}$ . The constant  $\sqrt{5}$  can be changed to a bigger one for  $\pi$ , but is otherwise optimal if  $\pi$  is replaced by the golden ratio  $\frac{\sqrt{5}+1}{2}$ .

The existence of infinitely many rational approximations as above gives us strictly increasing sequences  $m_k, n_k$  of natural numbers, such that  $\left| \frac{n_k}{m_k} - \pi \right| \leq \frac{1}{\sqrt{5}m_k^2}$ . This implies  $|n_k - m_k\pi| \leq \frac{1}{m_k}$  and  $\lim_{k \rightarrow \infty} (n_k - m_k\pi) = 0$ . Then the continuity of  $\sin x$  at 0 (see Section 2.2) implies  $\lim_{k \rightarrow \infty} \sin n_k = \lim_{k \rightarrow \infty} \sin(n_k - m_k\pi) = 0$ .

Exercise 1.81 gives a vast generalization of the example. It shows that  $\text{LIM}\{\sin n\} = [-1, 1]$ .

**Exercise 1.45.** For sequences in Exercise 1.14, find the sets of limits.

**Exercise 1.46.** Suppose the sequence  $z_n$  is obtained by combining two sequences  $x_n$  and  $y_n$  together. Prove that  $\text{LIM}\{z_n\} = \text{LIM}\{x_n\} \cup \text{LIM}\{y_n\}$ .

**Exercise 1.47.** Suppose  $l_k \in \text{LIM}\{x_n\}$  and  $\lim_{k \rightarrow \infty} l_k = l$ . Prove that  $l \in \text{LIM}\{x_n\}$ . In other words, the limit of limits is a limit.

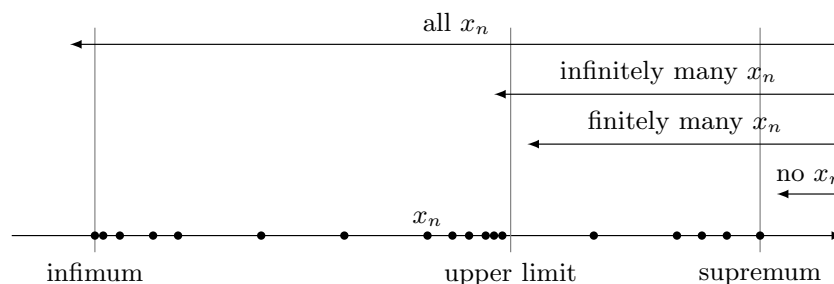
## Upper and Lower Limits

The supremum of  $\text{LIM}\{x_n\}$  is called the *upper limit* and denoted  $\overline{\lim}_{n \rightarrow \infty} x_n$ . The infimum of  $\text{LIM}\{x_n\}$  is called the *lower limit* and denoted  $\underline{\lim}_{n \rightarrow \infty} x_n$ . For example, for the sequence in Example 1.5.4, the upper limit is  $\max\{l_1, l_2, l_3\}$  and the lower limit is  $\min\{l_1, l_2, l_3\}$ .

The following characterizes the upper limit. The lower limit can be similarly characterized.

**Proposition 1.5.4.** Suppose  $x_n$  is a bounded sequence and  $l$  is a number.

1. If  $l > \overline{\lim}_{n \rightarrow \infty} x_n$ , then there are only finitely many  $x_n > l$ .
2. If  $l < \overline{\lim}_{n \rightarrow \infty} x_n$ , then there are infinitely many  $x_n > l$ .



**Figure 1.5.1.** Supremum, infimum, and upper limit.

The characterization provides the following picture for the upper limit. Let us start with a big  $l$  and move downwards. As  $l$  decrease, the number of terms  $x_n > l$  increases. If  $l$  is an upper bound of the sequence, then this number is zero. When  $l$  is lowered to no longer be an upper bound, some terms in the sequence will be bigger than  $l$ . The number may be finite at the beginning. But there will be a threshold, such that if  $l$  is below the threshold, then the number of  $x_n > l$  becomes infinite. The reason for the existence of such a threshold is that when  $l$  is so low to become a lower bound, then all (in particular, infinitely many)  $x_n > l$ .

This threshold is the upper limit.

*Proof.* The first statement is the same as the following: If there are infinitely many  $x_n > l$ , then  $l \leq \overline{\lim}_{n \rightarrow \infty} x_n$ . We will prove this equivalent statement.

Since there are infinitely many  $x_n > l$ , we can find a subsequence  $x_{n_k}$  satisfying  $x_{n_k} > l$ . By Bolzano-Weierstrass Theorem, the subsequence has a further convergent subsequence  $x_{n_{k_p}}$ . Then

$$\overline{\lim}_{n \rightarrow \infty} x_n = \sup \text{LIM}\{x_n\} \geq \lim x_{n_{k_p}} \geq l,$$

where the first inequality is the definition of  $\text{LIM}\{x_n\}$  and the second inequality follows from  $x_{n_k} > l$ .

For the second statement, let  $l < \overline{\lim}_{n \rightarrow \infty} x_n$ . By the definition of the upper limit, we have  $l < \lim x_{n_k}$  for some converging subsequence  $x_{n_k}$ . By Proposition 1.2.4, all the terms in the subsequence except finitely many will be bigger than  $l$ . Thus we find infinitely many  $x_n > l$ .  $\square$

**Proposition 1.5.5.** *The upper and lower limits are limits of convergent subsequences. Moreover, the sequence converges if and only if the upper and lower limits are equal.*

The first conclusion is  $\overline{\lim}_{n \rightarrow \infty} x_n, \underline{\lim}_{n \rightarrow \infty} x_n \in \text{LIM}\{x_n\}$ . In the second conclusion, the equality  $\overline{\lim}_{n \rightarrow \infty} x_n = \underline{\lim}_{n \rightarrow \infty} x_n = l$  means  $\text{LIM}\{x_n\} = \{l\}$ , which basically says that all convergent subsequences have the same limit.

*Proof.* Denote  $l = \overline{\lim} x_n$ . For any  $\epsilon > 0$ , we have  $l + \epsilon > \overline{\lim} x_n$  and  $l - \epsilon < \overline{\lim} x_n$ . By Proposition 1.5.4, there are only finitely many  $x_n > l + \epsilon$  and infinitely many  $x_n > l - \epsilon$ . Therefore there are infinitely many  $x_n$  satisfying  $l + \epsilon \geq x_n > l - \epsilon$ . Thus for any  $\epsilon > 0$ , there are infinitely many  $x_n$  satisfying  $|x_n - l| \leq \epsilon$ . By Proposition 1.5.3,  $l$  is the limit of a convergent subsequence.

For the second part, Proposition 1.2.2 says that if  $x_n$  converges to  $l$ , then  $\text{LIM}\{x_n\} = \{l\}$ , so that  $\overline{\lim} x_n = \underline{\lim} x_n = l$ . Conversely, suppose  $\overline{\lim} x_n = \underline{\lim} x_n = l$ . Then for any  $\epsilon > 0$ , we apply Proposition 1.5.4 to  $l + \epsilon > \overline{\lim} x_n$  and find only finitely many  $x_n > l + \epsilon$ . We also apply the similar characterization of the lower limit to  $l - \epsilon < \underline{\lim} x_n$  and find also only finitely many  $x_n < l - \epsilon$ . Thus  $|x_n - l| \leq \epsilon$  holds for all but finitely many  $x_n$ . If  $N$  is the biggest index for those  $x_n$  that do not satisfy  $|x_n - l| \leq \epsilon$ , then we get  $|x_n - l| \leq \epsilon$  for all  $n > N$ . This proves that  $x_n$  converges to  $l$ .  $\square$

**Exercise 1.48.** Find the upper and lower limits of bounded sequences in Exercise 1.14.

**Exercise 1.49.** Prove the properties of upper and lower limits.

1.  $\overline{\lim}_{n \rightarrow \infty} (-x_n) = -\underline{\lim}_{n \rightarrow \infty} x_n$ .
2.  $\overline{\lim}_{n \rightarrow \infty} x_n + \overline{\lim}_{n \rightarrow \infty} y_n \geq \overline{\lim}_{n \rightarrow \infty} (x_n + y_n) \geq \underline{\lim}_{n \rightarrow \infty} x_n + \overline{\lim}_{n \rightarrow \infty} y_n$ .
3. If  $x_n > 0$ , then  $\overline{\lim}_{n \rightarrow \infty} \frac{1}{x_n} = \frac{1}{\underline{\lim}_{n \rightarrow \infty} x_n}$ .
4. If  $x_n \geq 0$  and  $y_n \geq 0$ , then  $\overline{\lim}_{n \rightarrow \infty} x_n \cdot \overline{\lim}_{n \rightarrow \infty} y_n \geq \overline{\lim}_{n \rightarrow \infty} (x_n y_n) \geq \underline{\lim}_{n \rightarrow \infty} x_n \cdot \underline{\lim}_{n \rightarrow \infty} y_n$ .
5. If  $x_n \geq y_n$ , then  $\overline{\lim}_{n \rightarrow \infty} x_n \geq \overline{\lim}_{n \rightarrow \infty} y_n$  and  $\underline{\lim}_{n \rightarrow \infty} x_n \geq \underline{\lim}_{n \rightarrow \infty} y_n$ .
6. If  $\overline{\lim}_{n \rightarrow \infty} x_n > \overline{\lim}_{n \rightarrow \infty} y_n$  or  $\underline{\lim}_{n \rightarrow \infty} x_n > \underline{\lim}_{n \rightarrow \infty} y_n$ , then  $x_n > y_n$  for infinitely many  $n$ .
7. If for any  $N$ , there are  $m, n > N$ , such that  $x_m \geq y_n$ , then  $\overline{\lim} x_n \geq \underline{\lim} y_n$ .
8. If  $\overline{\lim} x_n > \underline{\lim} y_n$ , then for any  $N$ , there are  $m, n > N$ , such that  $x_m > y_n$ .

**Exercise 1.50.** Suppose the sequence  $z_n$  is obtained by combining two sequences  $x_n$  and  $y_n$  together. Prove that  $\overline{\lim} z_n = \max\{\overline{\lim} x_n, \overline{\lim} y_n\}$ .

**Exercise 1.51.** Prove that if  $\overline{\lim}_{n \rightarrow \infty} \left| \frac{x_{n+1}}{x_n} \right| < 1$ , then  $\lim_{n \rightarrow \infty} x_n = 0$ . Prove that if  $\underline{\lim}_{n \rightarrow \infty} \left| \frac{x_{n+1}}{x_n} \right| > 1$ , then  $\lim_{n \rightarrow \infty} x_n = \infty$ .

**Exercise 1.52.** Prove that the upper limit  $l$  of a bounded sequence  $x_n$  is characterized by the following two properties.

1.  $l$  is the limit of a convergent subsequence.
2. For any  $\epsilon > 0$ , there is  $N$ , such that  $x_n < l + \epsilon$  for any  $n > N$ .

The characterization may be compared with the one for the supremum in Exercise 1.33.

**Exercise 1.53.** For a Cauchy sequence, prove that any two converging subsequences have the same limit. Then by Proposition 1.5.5, the Cauchy sequence converges. This gives an alternative proof of the Cauchy criterion (Theorem 1.5.2).

**Exercise 1.54.** For a bounded sequence  $x_n$ , prove that

$$\begin{aligned}\overline{\lim}_{n \rightarrow \infty} x_n &= \lim_{n \rightarrow \infty} \sup \{x_n, x_{n+1}, x_{n+2}, \dots\}, \\ \underline{\lim}_{n \rightarrow \infty} x_n &= \lim_{n \rightarrow \infty} \inf \{x_n, x_{n+1}, x_{n+2}, \dots\}.\end{aligned}$$

## Heine-Borel Theorem

The Bolzano-Weierstrass Theorem has a set theoretical version that plays a crucial role in the point set topology. The property will not be needed for the analysis of single variable functions but will be useful for multivariable functions. The other reason for including the result here is that the proof is very similar to the proof of Bolzano-Weierstrass Theorem.

A set  $X$  of numbers is *closed* if  $x_n \in X$  and  $\lim_{n \rightarrow \infty} x_n = l$  implies  $l \in X$ . Intuitively, this means that one cannot escape  $X$  by taking limits. For example, the order rule (Proposition 1.2.4) says that closed intervals  $[a, b]$  are closed sets. In modern topological language, the following theorem says that bounded and closed sets of numbers are *compact*.

**Theorem 1.5.6** (Heine<sup>5</sup>-Borel<sup>6</sup> Theorem). *Suppose  $X$  is a bounded and closed set of numbers. Suppose  $\{(a_i, b_i)\}$  is a collection of open intervals such that  $X \subset \cup(a_i, b_i)$ . Then  $X \subset (a_{i_1}, b_{i_1}) \cup (a_{i_2}, b_{i_2}) \cup \dots \cup (a_{i_n}, b_{i_n})$  for finitely many intervals in the collection.*

We say  $\mathcal{U} = \{(a_i, b_i)\}$  is an *open cover* of  $X$  when  $X \subset \cup(a_i, b_i)$ . The theorem says that if  $X$  is bounded and closed, then any cover of  $X$  by open intervals has a

<sup>5</sup>Heinrich Eduard Heine, born March 15, 1821 in Berlin, Germany, died October 21, 1881 in Halle, Germany. In addition to the Heine-Borel theorem, Heine introduced the idea of uniform continuity.

<sup>6</sup>Félix Edouard Justin Émile Borel, born January 7, 1871 in Saint-Affrique, France, died February 3, 1956 in Paris France. Borel's measure theory was the beginning of the modern theory of functions of a real variable. He was French Minister of the Navy from 1925 to 1940.

finite *subcover*.

*Proof.* The bounded set  $X$  is contained in a bounded and closed interval  $[\alpha, \beta]$ . Suppose  $X$  cannot be covered by finitely many open intervals in  $\mathcal{U} = \{(a_i, b_i)\}$ .

Similar to the proof of Bolzano-Weierstrass Theorem, we divide  $[\alpha, \beta]$  into two halves  $I' = \left[\alpha, \frac{\alpha + \beta}{2}\right]$  and  $I'' = \left[\frac{\alpha + \beta}{2}, \beta\right]$ . Then either  $X \cap I'$  or  $X \cap I''$  cannot be covered by finitely many open intervals in  $\mathcal{U}$ . We denote the corresponding interval by  $I_1 = [\alpha_1, \beta_1]$  and denote  $X_1 = X \cap I_1$ .

Further divide  $I_1$  into  $I'_1 = \left[\alpha_1, \frac{\alpha_1 + \beta_1}{2}\right]$  and  $I''_1 = \left[\frac{\alpha_1 + \beta_1}{2}, \beta_1\right]$ . Again either  $X_1 \cap I'_1$  or  $X_1 \cap I''_1$  cannot be covered by finitely many open intervals in  $\mathcal{U}$ . We denote the corresponding interval by  $I_2 = [\alpha_2, \beta_2]$  and denote  $X_2 = X \cap I_2$ . Keep going, we get a sequence of intervals

$$I = [\alpha, \beta] \supset I_1 = [\alpha_1, \beta_1] \supset I_2 = [\alpha_2, \beta_2] \supset \cdots \supset I_n = [\alpha_n, \beta_n] \supset \cdots$$

with  $I_n$  having length  $\beta_n - \alpha_n = \frac{\beta - \alpha}{2^n}$ . Moreover,  $X_n = X \cap I_n$  cannot be covered by finitely many open intervals in  $\mathcal{U}$ . This implies that  $X_n$  is not empty, so that we can pick  $x_n \in X_n$ .

As argued in the proof of Bolzano-Weierstrass Theorem, we have converging limit  $l = \lim_{n \rightarrow \infty} \alpha_n = \lim_{n \rightarrow \infty} \beta_n$ . By  $\alpha_n \leq x_n \leq \beta_n$  and the sandwich rule, we get  $l = \lim_{n \rightarrow \infty} x_n$ . Then by the assumption that  $X$  is closed, we get  $l \in X$ .

Since  $X \subset \cup(a_i, b_i)$ , we have  $l \in (a_{i_0}, b_{i_0})$  for some interval  $(a_{i_0}, b_{i_0}) \in \mathcal{U}$ . Then by  $l = \lim_{n \rightarrow \infty} \alpha_n = \lim_{n \rightarrow \infty} \beta_n$ , we have  $X_n \subset I_n = [\alpha_n, \beta_n] \subset (a_{i_0}, b_{i_0})$  for sufficiently big  $n$ . In particular,  $X_n$  can be covered by one open interval in  $\mathcal{U}$ . The contradiction shows that  $X$  must be covered by finitely many open intervals from  $\mathcal{U}$ .  $\square$

**Exercise 1.55.** Find a collection  $\mathcal{U} = \{(a_i, b_i)\}$  that covers  $(0, 1]$ , but  $(0, 1]$  cannot be covered by finitely many intervals in  $\mathcal{U}$ . Find similar counterexample for  $[0, +\infty)$  in place of  $(0, 1]$ .

**Exercise 1.56 (Lebesgue<sup>7</sup>).** Suppose  $[\alpha, \beta]$  is covered by a collection  $\mathcal{U} = \{(a_i, b_i)\}$ . Denote

$$X = \{x \in [\alpha, \beta] : [\alpha, x] \text{ is covered by finitely many intervals in } \mathcal{U}\}.$$

1. Prove that  $\sup X \in X$ .
2. Prove that if  $x \in X$  and  $x < \beta$ , then  $x + \delta \in X$  for some  $\delta > 0$ .
3. Prove that  $\sup X = \beta$ .

This proves Heine-Borel Theorem for bounded and closed intervals.

<sup>7</sup>Henri Léon Lebesgue, born 1875 in Beauvais (France), died 1941 in Paris (France). His 1901 paper "Sur une généralisation de l'intégrale définie" introduced the concept of measure and revolutionized the integral calculus. He also made major contributions in other areas of mathematics, including topology, potential theory, the Dirichlet problem, the calculus of variations, set theory, the theory of surface area and dimension theory.



**Exercise 1.57.** Prove Heine-Borel Theorem for a bounded and closed set  $X$  in the following way. Suppose  $X$  is covered by a collection  $\mathcal{U} = \{(a_i, b_i)\}$ .

1. Prove that there is  $\delta > 0$ , such that for any  $x \in X$ ,  $(x - \delta, x + \delta) \subset (a_i, b_i)$  for some  $(a_i, b_i) \in \mathcal{U}$ .
2. Use the boundedness of  $X$  to find finitely many numbers  $c_1, c_2, \dots, c_n$ , such that  $X \subset (c_1, c_1 + \delta) \cup (c_2, c_2 + \delta) \cup \dots \cup (c_n, c_n + \delta)$ .
3. Prove that if  $X \cap (c_j, c_j + \delta) \neq \emptyset$ , then  $(c_j, c_j + \delta) \subset (a_i, b_i)$  for some  $(a_i, b_i) \in \mathcal{U}$ .
4. Prove that  $X$  is covered by no more than  $n$  open intervals in  $\mathcal{U}$ .

## 1.6 Additional Exercise

### Ratio Rule

**Exercise 1.58.** Suppose  $\left| \frac{x_{n+1}}{x_n} \right| \leq \left| \frac{y_{n+1}}{y_n} \right|$ .

1. Prove that  $|x_n| \leq c|y_n|$  for some constant  $c$ .
2. Prove that  $\lim_{n \rightarrow \infty} y_n = 0$  implies  $\lim_{n \rightarrow \infty} x_n = 0$ .
3. Prove that  $\lim_{n \rightarrow \infty} x_n = \infty$  implies  $\lim_{n \rightarrow \infty} y_n = \infty$ .

Note that in order to get the limit, the comparison only needs to hold for sufficiently big  $n$ .

**Exercise 1.59.** Suppose  $\lim_{n \rightarrow \infty} \frac{x_{n+1}}{x_n} = l$ . What can you say about  $\lim_{n \rightarrow \infty} x_n$  by looking at the value of  $l$ ?

**Exercise 1.60.** Use the ratio rule to get (1.1.5) and  $\lim_{n \rightarrow \infty} \frac{(n!)^2 a^n}{(2n)!}$ .

### Power Rule

The power rule says that if  $\lim_{n \rightarrow \infty} x_n = l > 0$ , then  $\lim_{n \rightarrow \infty} x_n^p = l^p$ . This is a special case of the exponential rule in Exercises 2.23 and 2.24.

**Exercise 1.61.** For integer  $p$ , show that the power rule is a special case of the arithmetic rule.

**Exercise 1.62.** Suppose  $x_n \geq 1$  and  $\lim_{n \rightarrow \infty} x_n = 1$ . Use the sandwich rule to prove that  $\lim_{n \rightarrow \infty} x_n^p = 1$  for any  $p$ . Moreover, show that the same is true if  $x_n \leq 1$ .

**Exercise 1.63.** Suppose  $\lim_{n \rightarrow \infty} x_n = 1$ . Use  $\min\{x_n, 1\} \leq x_n \leq \max\{x_n, 1\}$ , Exercise 1.18 and the sandwich rule to prove that  $\lim_{n \rightarrow \infty} x_n^p = 1$ .

**Exercise 1.64.** Prove the power rule in general.

### Average Rule

For a sequence  $x_n$ , the average sequence is  $y_n = \frac{x_1 + x_2 + \dots + x_n}{n}$ .

Exercise 1.65. Prove that if  $|x_n - l| < \epsilon$  for  $n > N$ , where  $N$  is a natural number, then

$$n > N \implies |y_n - l| < \frac{|x_1| + |x_2| + \cdots + |x_N| + N|l|}{n} + \epsilon.$$

Exercise 1.66. Prove that if  $\lim_{n \rightarrow \infty} x_n = l$ , then  $\lim_{n \rightarrow \infty} y_n = l$ .

Exercise 1.67. If  $\lim_{n \rightarrow \infty} x_n = \infty$ , can you conclude  $\lim_{n \rightarrow \infty} y_n = \infty$ ? What about  $+\infty$ ?

Exercise 1.68. Find suitable condition on a sequence  $a_n$  of positive numbers, such that  $\lim_{n \rightarrow \infty} x_n = l$  implies  $\lim_{n \rightarrow \infty} \frac{a_1 x_1 + a_2 x_2 + \cdots + a_n x_n}{a_1 + a_2 + \cdots + a_n} = l$ .

### Extended Supremum and Extended Upper Limit

Exercise 1.69. Extend the number system by including the “infinite numbers”  $+\infty$ ,  $-\infty$  and introduce the order  $-\infty < x < +\infty$  for any real number  $x$ . Then for any nonempty set  $X$  of real numbers and possibly  $+\infty$  or  $-\infty$ , we have  $\sup X$  and  $\inf X$  similarly defined. Prove that there are exactly three possibilities for  $\sup X$ .

1. If  $X$  has no finite number upper bound or  $+\infty \in X$ , then  $\sup X = +\infty$ .
2. If  $X$  has a finite number upper bound and contains at least one finite real number, then  $\sup X$  is a finite real number.
3. If  $X = \{-\infty\}$ , then  $\sup X = -\infty$ .

Write down the similar statements for  $\inf X$ .

Exercise 1.70. For a not necessarily bounded sequence  $x_n$ , extend the definition of  $\text{LIM}\{x_n\}$  by adding  $+\infty$  if there is a subsequence diverging to  $+\infty$ , and adding  $-\infty$  if there is a subsequence diverging to  $-\infty$ . Define the upper and lower limits as the supremum and infimum of  $\text{LIM}\{x_n\}$ , using the extension of the concepts in Exercise 1.69. Prove the following extensions of Proposition 1.5.5.

1. A sequence with no upper bound must have a subsequence diverging to  $+\infty$ . This means  $\overline{\lim}_{n \rightarrow \infty} x_n = +\infty$ .
2. If there is no subsequence with finite limit and no subsequence diverging to  $-\infty$ , then the whole sequence diverges to  $+\infty$ .

### Supremum and Infimum in Ordered Set

Recall that an order on a set is a relation  $x < y$  between pairs of elements satisfying the transitivity and the exclusivity. The concepts of upper bound, lower bound, supremum and infimum can be defined for subsets of an ordered set in a way similar to numbers.

Exercise 1.71. Provide a characterization of the supremum similar to numbers.

Exercise 1.72. Prove that the supremum, if exists, must be unique.

Exercise 1.73. An order is defined for all subsets of the plane  $\mathbb{R}^2$  by  $A \leq B$  if  $A$  is contained in  $B$ . Let  $R$  be the set of all rectangles centered at the origin and with circumference 1. Find the supremum and infimum of  $R$ .

**The Limits of the Sequence  $\sin n$** 

In Example 1.5.5, we used a theorem from number theory to prove that a subsequence of  $\sin n$  converges to 0. Here is a more general approach that proves that any number in  $[-1, 1]$  is the limit of a converging subsequence of  $\sin n$ .

The points on the unit circle  $S^1$  are described by the angles  $\theta$ . Note that the same point also corresponds to  $\theta + 2m\pi$  for integers  $m$ . Now for any angle  $\alpha$ , the rotation by angle  $\alpha$  is the map  $R_\alpha: S^1 \rightarrow S^1$  that takes the angle  $\theta$  to  $\theta + \alpha$ . If we apply the rotation  $n$  times, we get  $R_\alpha^n = R_{n\alpha}: S^1 \rightarrow S^1$  that takes the angle  $\theta$  to  $\theta + n\alpha$ . The rotation also has inverse  $R_\alpha^{-1} = R_{-\alpha}: S^1 \rightarrow S^1$  that takes the angle  $\theta$  to  $\theta - \alpha$ .

An interval  $(\theta_1, \theta_2)$  on the circle  $S^1$  consists of all points corresponding to angles  $\theta_1 < \theta < \theta_2$ . The interval may be also described as  $(\theta_1 + 2m\pi, \theta_2 + 2m\pi)$  and has length  $\theta_2 - \theta_1$ . A key property of the rotation is the preservation of angle length. In other words, the length of  $R_\alpha(\theta_1, \theta_2) = (\theta_1 + \alpha, \theta_2 + \alpha)$  is the same as the length of  $(\theta_1, \theta_2)$ .

We fix an angle  $\theta$  and also use  $\theta$  to denote the corresponding point on the circle. We also fix a rotation angle  $\alpha$  and consider all the points

$$X = \{R_\alpha^n(\theta) = \theta + n\alpha : n \in \mathbb{Z}\}$$

obtained by rotating  $\theta$  by angle  $\alpha$  repeatedly.

**Exercise 1.74.** Prove that if an interval  $(\theta_1, \theta_2)$  on the circle  $S^1$  does not contain points in  $X$ , then for any  $n$ ,  $R_\alpha^n(\theta_1, \theta_2)$  does not contain points in  $X$ .

**Exercise 1.75.** Suppose  $(\theta_1, \theta_2)$  is a maximal open interval that does not contain points in  $X$ . In other words, any bigger open interval will contain some points in  $X$ . Prove that for any  $n$ , either  $R_\alpha^n(\theta_1, \theta_2)$  is disjoint from  $(\theta_1, \theta_2)$  or  $R_\alpha^n(\theta_1, \theta_2) = (\theta_1, \theta_2)$ .

**Exercise 1.76.** Suppose there is an open interval containing no points in  $X$ . Prove that there is a maximal open interval  $(\theta_1, \theta_2)$  containing no points in  $X$ . Moreover, we have  $R_\alpha^n(\theta_1, \theta_2) = (\theta_1, \theta_2)$  for some natural number  $n$ .

**Exercise 1.77.** Prove that  $R_\alpha^n(\theta_1, \theta_2) = (\theta_1, \theta_2)$  for some natural number  $n$  if and only if  $\alpha$  is a rational multiple of  $\pi$ .

**Exercise 1.78.** Prove that if  $\alpha$  is not a rational multiple of  $\pi$ , then every open interval contains some point in  $X$ .

**Exercise 1.79.** Prove that if every open interval contains some point in  $X$ , then for any angle  $l$ , there is a sequence  $n_k$  of integers that diverge to infinity, such that the sequence  $R_\alpha^{n_k}(\theta) = \theta + n_k\alpha$  on the circle converges to  $l$ . Then interpret the result as  $\lim_{k \rightarrow \infty} |\theta + n_k\alpha - 2m_k\pi| = 0$  for another sequence of integers  $m_k$ .

**Exercise 1.80.** Prove that if  $\alpha$  is not a rational multiple of  $\pi$ , then there are sequences of natural numbers  $m_k, n_k$  diverging to  $+\infty$ , such that  $\lim_{k \rightarrow \infty} |n_k\alpha - 2m_k\pi| = 0$ . Then use the result to improve the  $n_k$  in Exercise 1.79 to be natural numbers.

**Exercise 1.81.** Suppose  $\theta$  is any number and  $\alpha$  is not a rational multiple of  $\pi$ . Prove that any number in  $[-1, 1]$  is the limit of a converging subsequence of  $\sin(\theta + n\alpha)$ .

### Alternative Proof of Bolzano-Weierstrass Theorem

We say a term  $x_n$  in a sequence has property  $P$  if there is  $M$ , such that  $m > M$  implies  $x_m > x_n$ . The property means that  $x_m > x_n$  for sufficiently big  $m$ .

**Exercise 1.82.** Suppose there are infinitely many terms in a sequence  $x_n$  with property  $P$ . Construct an increasing subsequence  $x_{n_k}$  in which each  $x_{n_k}$  has property  $P$ . Moreover, in case  $x_n$  is bounded, prove that  $x_{n_k}$  converges to  $\lim x_n$ .

**Exercise 1.83.** Suppose there are only finitely many terms in a sequence  $x_n$  with property  $P$ . Construct a decreasing subsequence  $x_{n_k}$ .

### Set Version of Bolzano-Weierstrass Theorem, Upper Limit and Lower Limit

Let  $X$  be a set of numbers. A number  $l$  is a *limit* of  $X$  if for any  $\epsilon > 0$ , there is  $x \in X$  satisfying  $0 < |x - l| < \epsilon$ . The definition is similar to the characterisation of the limit of a subsequence in Proposition 1.5.3. Therefore the limit of a set is the analogue of limit of converging subsequences. We may similarly denote all the limits of the set by  $\text{LIM} X$  (the common notation in topology is  $X'$ , called *derived set*).

**Exercise 1.84.** Prove that  $l$  is a limit of  $X$  if and only if  $l$  is the limit of a *non repetitive* (i.e., no two terms are equal) sequence in  $X$ . This is the analogue of Proposition 1.5.3.

**Exercise 1.85.** What does it mean for a number not to be a limit? Explain that a finite set has no limit.

**Exercise 1.86.** Prove the set version of Bolzano-Weierstrass Theorem: Any infinite, bounded and *closed* set of numbers has an accumulation point. Here we say  $X$  is closed if  $\lim x_n = l$  and  $x_n \in X$  imply  $l \in X$ .

**Exercise 1.87.** Use the set version of Bolzano-Weierstrass Theorem to prove Theorem 1.5.1, the sequence version.

**Exercise 1.88.** Use  $\text{LIM} X$  to define the upper limit  $\overline{\lim} X$  and the lower limit  $\underline{\lim} X$  of a set. Then establish the analogue of Proposition 1.5.4 that characterises the two limits.

**Exercise 1.89.** Prove that  $\overline{\lim} X$  and  $\underline{\lim} X$  are also limits of  $X$ . Moreover, explain what happens when  $\overline{\lim} X = \underline{\lim} X$ . The problem is the analogue of Proposition 1.5.5.

**Exercise 1.90.** Try to extend Exercises 1.49, 1.50, 1.54 to infinite and bounded sets of numbers.

## Chapter 2

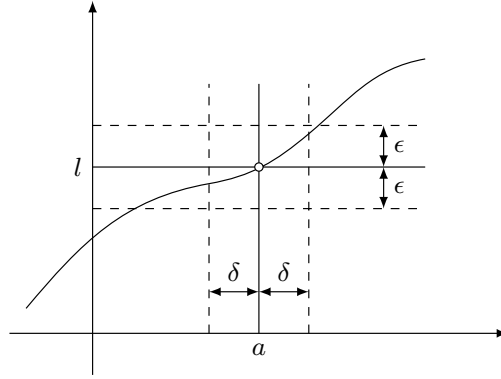
# Limit of Function

## 2.1 Definition

For a function  $f(x)$  defined near  $a$  (but not necessarily at  $a$ ), we may consider its behavior as  $x$  approaches  $a$ .

**Definition 2.1.1.** A function  $f(x)$  defined near  $a$  has *limit*  $l$  at  $a$ , and denoted  $\lim_{x \rightarrow a} f(x) = l$ , if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$0 < |x - a| < \delta \implies |f(x) - l| < \epsilon. \quad (2.1.1)$$



**Figure 2.1.1.**  $0 < |x - a| < \delta$  implies  $|f(x) - l| < \epsilon$ .

The definition says

$$x \rightarrow a, x \neq a \implies f(x) \rightarrow l.$$

Similar to the sequence limit, the smallness  $\epsilon$  for  $|f(x) - l|$  is *arbitrarily* given, while the size  $\delta$  for  $|x - a|$  is to be found *after*  $\epsilon$  is given. Thus the choice of  $\delta$  usually depends on  $\epsilon$  and is often expressed as a function of  $\epsilon$ . Moreover, since the limit is about how close the numbers are, only small  $\epsilon$  and  $\delta$  need to be considered.

### Variations of Function Limit

In the definition of function limit,  $x$  may approach  $a$  from the right (i.e.,  $x > a$ ) or from the left (i.e.,  $x < a$ ). The two approaches may be treated separately, leading to *one sided limits*.

**Definition 2.1.2.** A function  $f(x)$  defined for  $x > a$  and near  $a$  has *right limit*  $l$  at  $a$ , and denoted  $\lim_{x \rightarrow a^+} f(x) = l$ , if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$0 < x - a < \delta \implies |f(x) - l| < \epsilon. \quad (2.1.2)$$

A function  $f(x)$  defined for  $x < a$  and near  $a$  has *left limit*  $l$  at  $a$ , and denoted  $\lim_{x \rightarrow a^-} f(x) = l$ , if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$-\delta < x - a < 0 \implies |f(x) - l| < \epsilon. \quad (2.1.3)$$

The one sided limits are often denoted as  $f(a^+) = \lim_{x \rightarrow a^+} f(x)$  and  $f(a^-) = \lim_{x \rightarrow a^-} f(x)$ . Moreover, the one sided limits and the usual (two sided) limit are related as follows.

**Proposition 2.1.3.**  $\lim_{x \rightarrow a} f(x) = l$  if and only if  $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^-} f(x) = l$ .

*Proof.* The two sided limit implies the two one sided limits, because (2.1.1) implies (2.1.2) and (2.1.3). Conversely, if  $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^-} f(x) = l$ , then for any  $\epsilon > 0$ , there are  $\delta_+, \delta_- > 0$ , such that

$$\begin{aligned} 0 < x - a < \delta_+ &\implies |f(x) - l| < \epsilon, \\ -\delta_- < x - a < 0 &\implies |f(x) - l| < \epsilon. \end{aligned}$$

Then for  $0 < |x - a| < \delta = \min\{\delta_+, \delta_-\}$ , we have either  $0 < x - a < \delta \leq \delta_+$  or  $-\delta_- \leq -\delta < x - a < 0$ . In either case, we get  $|f(x) - l| < \epsilon$ .  $\square$

We may also define the function limit when  $x$  gets very big.

**Definition 2.1.4.** A function  $f(x)$  has limit  $l$  at  $\infty$ , denoted  $\lim_{x \rightarrow \infty} f(x) = l$ , if for any  $\epsilon > 0$ , there is  $N$ , such that

$$|x| > N \implies |f(x) - l| < \epsilon.$$

The limit at infinity can also be split into the limits  $f(+\infty) = \lim_{x \rightarrow +\infty} f(x)$ ,  $f(-\infty) = \lim_{x \rightarrow -\infty} f(x)$  at positive and negative infinities. Proposition 2.1.3 also holds for the limit at infinity.

The divergence to infinity can also be defined for functions.

**Definition 2.1.5.** A function  $f(x)$  diverges to *infinity* at  $a$ , denoted  $\lim_{x \rightarrow a} f(x) = \infty$ , if for any  $b$ , there is  $\delta > 0$ , such that

$$0 < |x - a| < \delta \implies |f(x)| > b.$$

The divergence to positive and negative infinities, denoted  $\lim_{x \rightarrow a} f(x) = +\infty$  and  $\lim_{x \rightarrow a} f(x) = -\infty$  respectively, can be similarly defined. Moreover, the divergence to infinity at the left of  $a$ , the right of  $a$ , or when  $a$  is various kinds of infinities, can also be similarly defined.

Similar to sequences, we know  $\lim_{x \rightarrow a} f(x) = \infty$  if and only if  $\lim_{x \rightarrow a} \frac{1}{f(x)} = 0$ , i.e., the reciprocal is an *infinitesimal*.

**Exercise 2.1.** Write down the rigorous definition of  $\lim_{x \rightarrow +\infty} f(x) = l$ ,  $\lim_{x \rightarrow a^+} f(x) = -\infty$ ,  $\lim_{x \rightarrow \infty} f(x) = +\infty$ .

**Exercise 2.2.** Suppose  $f(x) \leq l \leq g(x)$  and  $\lim_{x \rightarrow a} (f(x) - g(x)) = 0$ . Prove that  $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = l$ . Extend Exercises 1.3 through 1.7 in similar way.

**Exercise 2.3.** Prove the following are equivalent definitions of  $\lim_{x \rightarrow a} f(x) = l$ .

1. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $0 < |x - a| \leq \delta$  implies  $|f(x) - l| < \epsilon$ .
2. For any  $c > \epsilon > 0$ , where  $c$  is some fixed number, there is  $\delta > 0$ , such that  $0 < |x - a| < \delta$  implies  $|f(x) - l| \leq \epsilon$ .
3. For any natural number  $n$ , there is  $\delta > 0$ , such that  $0 < |x - a| < \delta$  implies  $|f(x) - l| \leq \frac{1}{n}$ .
4. For any  $1 > \epsilon > 0$ , there is  $\delta > 0$ , such that  $0 < |x - a| < \delta$  implies  $|f(x) - l| < \frac{\epsilon}{1 - \epsilon}$ .

**Exercise 2.4.** Which are equivalent to the definition of  $\lim_{x \rightarrow a} f(x) = l$ ?

1. For  $\epsilon = 0.001$ , we have  $\delta = 0.01$ , such that  $0 < |x - a| \leq \delta$  implies  $|f(x) - l| < \epsilon$ .
2. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $|x - a| < \delta$  implies  $|f(x) - l| < \epsilon$ .
3. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $0 < |x - a| < \delta$  implies  $0 < |f(x) - l| < \epsilon$ .
4. For any  $0.001 \geq \epsilon > 0$ , there is  $\delta > 0$ , such that  $0 < |x - a| \leq 2\delta$  implies  $|f(x) - l| \leq \epsilon$ .
5. For any  $\epsilon > 0.001$ , there is  $\delta > 0$ , such that  $0 < |x - a| \leq 2\delta$  implies  $|f(x) - l| \leq \epsilon$ .
6. For any  $\epsilon > 0$ , there is a rational number  $\delta > 0$ , such that  $0 < |x - a| < \delta$  implies  $|f(x) - l| < \epsilon$ .
7. For any  $\epsilon > 0$ , there is a natural number  $N$ , such that  $0 < |x - a| < \frac{1}{N}$  implies  $|f(x) - l| < \epsilon$ .
8. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $0 < |x - a| < \delta$  implies  $|f(x) - l| < \epsilon^2$ .
9. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $0 < |x - a| < \delta$  implies  $|f(x) - l| < \epsilon^2 + 1$ .
10. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $0 < |x - a| < \delta + 1$  implies  $|f(x) - l| < \epsilon^2$ .

## Basic Properties of the Function Limit

The properties of limit of sequences can be extended to functions.

**Proposition 2.1.6.** *The limit of functions has the following properties.*

1. *Boundedness:* A function convergent at  $a$  is bounded near  $a$ .
2. *Arithmetic:* Suppose  $\lim_{x \rightarrow a} f(x) = l$  and  $\lim_{x \rightarrow a} g(x) = k$ . Then

$$\lim_{x \rightarrow a} (f(x) + g(x)) = l + k, \quad \lim_{x \rightarrow a} f(x)g(x) = lk, \quad \lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{l}{k},$$

where  $g(x) \neq 0$  and  $k \neq 0$  are assumed in the third equality.

3. *Order:* Suppose  $\lim_{x \rightarrow a} f(x) = l$  and  $\lim_{x \rightarrow a} g(x) = k$ . If  $f(x) \geq g(x)$  for  $x$  near  $a$ , then  $l \geq k$ . Conversely, if  $l > k$ , then  $f(x) > g(x)$  for  $x$  near  $a$ .
4. *Sandwich:* Suppose  $f(x) \leq g(x) \leq h(x)$  for  $x$  near  $a$  and  $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} h(x) = l$ . Then  $\lim_{x \rightarrow a} g(x) = l$ .



5. *Composition:* Suppose  $\lim_{x \rightarrow a} f(x) = b$ ,  $\lim_{y \rightarrow b} g(y) = c$ . If  $f(x) \neq b$  for  $x$  near  $a$ , or  $g(b) = c$ , then  $\lim_{x \rightarrow a} g(f(x)) = c$ .

When we say something happens *near*  $a$ , we mean that there is  $\delta > 0$ , such that it happens for  $x$  satisfying  $0 < |x - a| < \delta$ . For example,  $f(x)$  is bounded near  $a$  if there is  $\delta > 0$  and  $B$ , such that  $0 < |x - a| < \delta$  implies  $|f(x)| < B$ .

For a composition

$$x \mapsto y = f(x) \mapsto z = g(y) = g(f(x)),$$

the composition rule says

$$\lim_{x \rightarrow a} y = b, \lim_{y \rightarrow b} z = c \implies \lim_{x \rightarrow a} z = c.$$

The left side means two implications

$$\begin{aligned} x \rightarrow a, x \neq a &\implies y \rightarrow b, \\ y \rightarrow b, y \neq b &\implies z \rightarrow c. \end{aligned}$$

However, to combine the two implications to get what we want

$$x \rightarrow a, x \neq a \implies z \rightarrow c,$$

the right side of the first implication needs to match the left side of the second implication. To make the match, we need to either modify the first implication to

$$x \rightarrow a, x \neq a \implies y \rightarrow b, y \neq b,$$

which is the first additional condition in the fifth statement of Proposition 2.1.6, or modify the second implication to

$$y \rightarrow b \implies z \rightarrow c,$$

which is the second additional condition.

*Proof.* The first four properties are parallel to the Propositions 1.2.1 through 1.2.5 for the sequence limit, and can be proved in the similar way.

Now we turn to the composition rule. For any  $\epsilon > 0$ , by  $\lim_{y \rightarrow b} g(y) = c$ , there is  $\mu > 0$ , such that

$$0 < |y - b| < \mu \implies |g(y) - c| < \epsilon. \quad (2.1.4)$$

For this  $\mu > 0$ , by  $\lim_{x \rightarrow a} f(x) = b$ , there is  $\delta > 0$ , such that

$$0 < |x - a| < \delta \implies |f(x) - b| < \mu. \quad (2.1.5)$$

If the additional condition  $f(x) \neq b$  for  $x$  near  $a$  is satisfied, then (2.1.5) becomes

$$0 < |x - a| < \delta \implies 0 < |f(x) - b| < \mu.$$

Combining this with (2.1.4) and taking  $y = f(x)$ , we get

$$0 < |x - a| < \delta \implies 1 < |f(x) - b| < \mu \implies |g(f(x)) - c| < \epsilon.$$

If the additional condition  $g(b) = c$  is satisfied, then (2.1.4) becomes

$$|y - b| < \mu \implies |g(y) - c| < \epsilon.$$

Combining this with (2.1.5) and taking  $y = f(x)$ , we get

$$0 < |x - a| < \delta \implies |f(x) - b| < \mu \implies |g(f(x)) - c| < \epsilon.$$

So under either additional condition, we proved  $\lim_{y \rightarrow a} g(f(x)) = c$ .  $\square$

Proposition 2.1.6 was stated for the two sided limit  $\lim_{x \rightarrow a} f(x) = l$  with finite  $a$  and  $l$  only. The properties also hold when  $a$  is replaced by  $a^+$ ,  $a^-$ ,  $\infty$ ,  $+\infty$  and  $-\infty$ .

What about the case that  $l$  is infinity? In general, all the valid arithmetic rules for sequences that involve infinities and infinitesimals, such as  $(+\infty) + (+\infty) = +\infty$ , are still valid for functions. However, as in the sequence case, the same care needs to be taken in applying the arithmetic rules to infinities and infinitesimals.

For the sandwich rule, we have  $f(x) \geq g(x)$  and  $\lim_{x \rightarrow a} g(x) = +\infty$  implying  $\lim_{x \rightarrow a} f(x) = +\infty$ . There is similar sandwich rule for  $l = -\infty$  but no sandwich rule for  $l = \infty$ .

For the composition rule,  $a, b, c$  can be practically any symbols. For example, if  $\lim_{x \rightarrow \infty} g(x) = b$ ,  $g(x) > b$  and  $\lim_{y \rightarrow b^+} f(y) = c$ , then  $\lim_{x \rightarrow \infty} f(g(x)) = c$ .

**Example 2.1.1.** The limit of power functions will be established in Section 2.2. Suppose we know  $\lim_{x \rightarrow 0} x^2 = 0$  and  $\lim_{x \rightarrow 0^+} \sqrt{x} = 0$ . We will argue that  $\lim_{x \rightarrow 0} f(x^2) = \lim_{x \rightarrow 0^+} f(x)$ .

If  $\lim_{x \rightarrow 0} f(x^2) = l$  converges, then consider the composition

$$x \mapsto y = \sqrt{x} \mapsto z = f(y^2) = f(x).$$

By  $\lim_{x \rightarrow 0^+} \sqrt{x} = 0$ ,  $\lim_{x \rightarrow 0} f(x^2) = l$ , and the fact that  $x > 0 \implies \sqrt{x} > 0$  (so the first additional condition is satisfied), we may apply the composition rule to get  $\lim_{x \rightarrow 0^+} f(x) = l$ .

Conversely, if  $\lim_{x \rightarrow 0^+} f(x) = l$  converges, then consider the composition

$$x \mapsto y = x^2 \mapsto z = f(y) = f(x^2).$$

By  $\lim_{x \rightarrow 0} x^2 = 0$ ,  $\lim_{x \rightarrow 0^+} f(x) = l$ , and the fact that  $x \neq 0 \implies x^2 > 0$ , we may apply the composition rule to get  $\lim_{x \rightarrow 0} f(x^2) = l$ .

**Exercise 2.5.** Prove the first four properties of Proposition 2.1.6.

**Exercise 2.6.** Assume that we already know the limit of power functions. Rewrite the limits as  $\lim_{x \rightarrow a} f(x)$  for suitable  $a$  and explain whether the limits are equivalent.

- |  |  |  |
|--|--|--|
| 1. $\lim_{x \rightarrow a^-} f(-x).$   | 4. $\lim_{x \rightarrow 0^+} f(x^2).$      | 7. $\lim_{x \rightarrow 0} f\left(\frac{1}{x}\right).$   |
| 2. $\lim_{x \rightarrow a^+} f(x+1).$  | 5. $\lim_{x \rightarrow 0} f((x+1)^3).$    |  |
| 3. $\lim_{x \rightarrow a^+} f(bx+c).$ | 6. $\lim_{x \rightarrow 0^+} f(\sqrt{x}).$ | 8. $\lim_{x \rightarrow 2^+} f\left(\frac{1}{x}\right).$ |

**Exercise 2.7.** Prove properties of the function limit.

1. If  $\lim_{x \rightarrow a} f(x) = l$  and  $\lim_{x \rightarrow a} g(x) = k$ , then  $\lim_{x \rightarrow a} \max\{f(x), g(x)\} = \max\{l, k\}$  and  $\lim_{x \rightarrow a} \min\{f(x), g(x)\} = \min\{l, k\}$ .
2. If  $\lim_{x \rightarrow a^+} f(x) = \infty$  and there are  $c > 0$  and  $\delta > 0$ , such that  $0 < x - a < \delta$  implies  $g(x) > c$ , then  $\lim_{x \rightarrow a^+} f(x)g(x) = \infty$ .
3. If  $\lim_{x \rightarrow a} g(x) = +\infty$  and  $\lim_{y \rightarrow +\infty} f(y) = c$ , then  $\lim_{x \rightarrow a} f(g(x)) = c$ .
4. If  $f(x) \leq g(x)$  and  $\lim_{x \rightarrow +\infty} g(x) = -\infty$ , then  $\lim_{x \rightarrow +\infty} f(x) = -\infty$ .

## Function Limit and Sequence Limit

A sequence can be considered as a function  $n \mapsto x_n$ . The composition of a function  $f(x)$  with a sequence

$$n \mapsto x = x_n \mapsto y = f(x) = f(x_n)$$

is the restriction of the function to the sequence.

**Proposition 2.1.7.** Suppose  $f(x)$  is a function defined near  $a$ . Then  $\lim_{x \rightarrow a} f(x) = l$  if and only if  $\lim_{n \rightarrow \infty} f(x_n) = l$  for any sequence  $x_n$  satisfying  $x_n \neq a$  and  $\lim_{n \rightarrow \infty} x_n = a$ .

*Proof.* We prove the case  $a$  and  $l$  are finite. The other cases are similar.

Suppose  $\lim_{x \rightarrow a} f(x) = l$ . Suppose  $x_n \neq a$  and  $\lim_{n \rightarrow \infty} x_n = a$ . For any  $\epsilon > 0$ , we can find  $\delta > 0$ , such that

$$0 < |x - a| < \delta \implies |f(x) - l| < \epsilon.$$

Then we can find  $N$ , such that

$$n > N \implies |x_n - a| < \delta.$$

The assumption  $x_n \neq a$  further implies

$$n > N \implies 0 < |x_n - a| < \delta.$$

Combining the two implications together, we have

$$n > N \implies 0 < |x_n - a| < \delta \implies |f(x_n) - l| < \epsilon.$$

This proves  $\lim_{n \rightarrow \infty} f(x_n) = l$ .

Conversely, assume  $\lim_{x \rightarrow a} f(x) \neq l$  (which means either the limit does not exist, or the limit exists but is not equal to  $l$ ). Then there is  $\epsilon > 0$ , such that for any

$\delta > 0$ , there is  $x$  satisfying  $0 < |x - a| < \delta$  and  $|f(x) - l| \geq \epsilon$ . Specifically, by choosing  $\delta = \frac{1}{n}$  for natural numbers  $n$ , we find a sequence  $x_n$  satisfying  $0 < |x_n - a| < \frac{1}{n}$  and  $|f(x_n) - l| \geq \epsilon$ . The first inequality implies  $x_n \neq a$  and  $\lim_{n \rightarrow \infty} x_n = a$ . The second inequality implies  $\lim_{n \rightarrow \infty} f(x_n) \neq l$ . This proves the converse.  $\square$

We remark that a subsequence can be considered as the composition of two sequences

$$k \mapsto n = n_k \mapsto x = x_{n_k}.$$

Then Proposition 1.2.2 about the convergence of subsequences can also be considered as a version of the composition rule.

**Example 2.1.2.** The *Dirichlet*<sup>8</sup> function is

$$D(x) = \begin{cases} 1, & \text{if } x \text{ is rational,} \\ 0, & \text{if } x \text{ is irrational.} \end{cases}$$

For any  $a$ , we can find a sequence  $x_n$  of rational numbers and a sequence  $y_n$  of irrational numbers converging to but not equal to  $a$  (for  $a = 0$ , take  $x_n = \frac{1}{n}$  and  $y_n = \frac{\sqrt{2}}{n}$ , for example). Then  $f(x_n) = 1$  and  $f(y_n) = 0$ , so that  $\lim_{n \rightarrow \infty} f(x_n) = 1$  and  $\lim_{n \rightarrow \infty} f(y_n) = 0$ . This implies that the Dirichlet function diverges everywhere.

**Exercise 2.8.** Prove that  $\lim_{x \rightarrow \infty} f(x) = +\infty$  if and only if  $\lim_{n \rightarrow \infty} f(x_n) = +\infty$  for any sequence  $x_n$  satisfying  $\lim_{n \rightarrow \infty} x_n = \infty$ .

**Exercise 2.9.** Prove that  $\lim_{x \rightarrow a} f(x)$  converges if and only if  $\lim_{n \rightarrow \infty} f(x_n)$  converges for any sequence  $x_n$  satisfying  $x_n \neq a$  and  $\lim_{n \rightarrow \infty} x_n = a$ .

**Exercise 2.10.** Prove that  $\lim_{x \rightarrow a^+} f(x) = l$  if and only if  $\lim_{n \rightarrow \infty} f(x_n) = l$  for any *strictly decreasing* sequence  $x_n$  converging to  $a$ . Moreover, state the similar criterion for  $\lim_{x \rightarrow +\infty} f(x) = l$ .

## Monotone function

A function is *increasing* if

$$x > y \implies f(x) \geq f(y).$$

It is *strictly increasing* if

$$x > y \implies f(x) > f(y).$$

The concepts of *decreasing* and *strictly decreasing* functions can be similarly defined. A function is *monotone* if it is either increasing or decreasing.

<sup>8</sup>Johann Peter Gustav Lejeune Dirichlet, born 1805 in Düren (French Empire, now Germany), died in Göttingen (Germany). He proved the famous Fermat's Last Theorem for the case  $n = 5$  in 1825. He made fundamental contributions to the analytic number theory, partial differential equation, and Fourier series. He introduced his famous function in 1829.

Proposition 1.4.4 can be extended to the one sided limit and the limit at signed infinities of monotone functions. The following is stated for the right limit, and can be proved in similar way. See Exercises 2.11 and 2.12.

**Proposition 2.1.8.** *Suppose  $f(x)$  is a monotone bounded function defined for  $x > a$  and  $x$  near  $a$ . Then  $\lim_{x \rightarrow a^+} f(x)$  converges.*

**Exercise 2.11.** Suppose  $f(x)$  is increasing on  $(a, b]$ .

1. If  $f(x)$  is bounded, prove that  $\lim_{x \rightarrow a^+} f(x)$  converges to  $\inf_{(a, b]} f(x)$ .
2. If  $f(x)$  is unbounded, prove that  $\lim_{x \rightarrow a^+} f(x) = -\infty$ .

**Exercise 2.12.** State the similar version of Exercise 2.11 for  $\lim_{x \rightarrow +\infty} f(x)$ .

## Cauchy Criterion

**Theorem 2.1.9 (Cauchy Criterion).** *The limit  $\lim_{x \rightarrow a} f(x)$  converges if and only if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that*

$$0 < |x - a| < \delta, 0 < |y - a| < \delta \implies |f(x) - f(y)| < \epsilon. \quad (2.1.6)$$

*Proof.* Suppose  $\lim_{x \rightarrow a} f(x) = l$ . For any  $\epsilon > 0$ , there is  $\delta$ , such that  $0 < |x - a| < \delta$  implies  $|f(x) - l| < \frac{\epsilon}{2}$ . Then  $0 < |x - a| < \delta$  and  $0 < |y - a| < \delta$  imply

$$|f(x) - f(y)| = |(f(x) - l) - (f(y) - l)| \leq |f(x) - l| + |f(y) - l| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Conversely, assume  $f(x)$  satisfies the Cauchy criterion. We first prove that  $f(x)$  converges on any sequence satisfying  $x_n \neq a$  and  $\lim_{n \rightarrow \infty} x_n = a$ . Then we prove that the convergence of “subsequence”  $f(x_n)$  forces the convergence of the whole “sequence”  $f(x)$ , just like the third step in the proof of Theorem 1.5.2.

By Theorem 1.5.2, to show the convergence of  $f(x_n)$ , we only need to show that it is a Cauchy sequence. For any  $\epsilon > 0$ , we find  $\delta > 0$  such that (2.1.6) holds. Then for this  $\delta > 0$ , there is  $N$ , such that

$$n > N \implies 0 < |x_n - a| < \delta,$$

where  $0 < |x_n - a|$  follows from the assumption  $|x_n - a| > 0$ . Combined with (2.1.6), we get

$$\begin{aligned} m, n > N &\implies 0 < |x_m - a| < \delta, 0 < |x_n - a| < \delta \\ &\implies |f(x_m) - f(x_n)| < \epsilon. \end{aligned}$$

This proves that  $f(x_n)$  is a Cauchy sequence and therefore converges to a limit  $l$ .

Next we prove that  $l$  is also the limit of the function at  $a$ . For any  $\epsilon > 0$ , we find  $\delta > 0$  such that (2.1.6) holds. Then we find one  $x_n$  satisfying  $0 < |x_n - a| < \delta$  and  $|f(x_n) - l| < \epsilon$ . Moreover, for any  $x$  satisfying  $0 < |x - a| < \delta$ , we may apply (2.1.6) to  $x$  and  $x_n$  to get  $|f(x) - f(x_n)| < \epsilon$ . Therefore

$$|f(x) - l| \leq |f(x_n) - l| + |f(x) - f(x_n)| < 2\epsilon.$$

This completes the proof that  $\lim_{x \rightarrow a} f(x) = l$ .  $\square$

We note that the strategy for extending the Cauchy criterion to function limit is to first restrict to a Cauchy (and therefore converging) subsequence, and then show that the Cauchy criterion extends the convergence of the subsequence to the convergence of the whole “sequence”. The strategy works in very general settings and gives the Cauchy criterion in very general setting. See Exercises 2.101 and 2.102.

**Exercise 2.13.** State and prove the Cauchy criterion for the convergence of  $\lim_{x \rightarrow a^+} f(x)$  and  $\lim_{x \rightarrow +\infty} f(x)$ .

**Exercise 2.14.** Suppose for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$a - \delta < x < a < y < a + \delta \implies |f(x) - f(y)| < \epsilon.$$

Prove that  $\lim_{x \rightarrow a} f(x)$  converges.

## 2.2 Basic Limit

This section is devoted to rigorously deriving the important and basic function limits.

### Polynomial and Rational Function

It is easy to see that  $\lim_{x \rightarrow a} c = c$  for a constant  $c$  and  $\lim_{x \rightarrow a} x = a$ . Then by repeatedly applying the arithmetic rule, we get the limit of a *polynomial*

$$\lim_{x \rightarrow a} (c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0) = c_n a^n + c_{n-1} a^{n-1} + \cdots + c_1 a + c_0.$$

Note that the limit is simply the value of the polynomial at  $a$ . More generally, a *rational function*

$$r(x) = \frac{c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0}{d_m x^m + d_{m-1} x^{m-1} + \cdots + d_1 x + d_0}$$

is a quotient of two polynomials. By the limit of polynomial and the arithmetic rule, we get

$$\lim_{x \rightarrow a} r(x) = \frac{c_n a^n + c_{n-1} a^{n-1} + \cdots + c_1 a + c_0}{d_m a^m + d_{m-1} a^{m-1} + \cdots + d_1 a + d_0} = r(a),$$

as long as the denominator is nonzero.

We also have  $\lim_{x \rightarrow \infty} c = c$  and  $\lim_{x \rightarrow \infty} \frac{1}{x} = 0$ . Then by the arithmetic rule,

we get

$$\lim_{x \rightarrow \infty} \frac{2x^5 + 10}{-3x + 1} = \lim_{x \rightarrow \infty} x^4 \frac{2 + 10\frac{1}{x}}{-3 + \frac{1}{x}} = (+\infty) \frac{2 + 10 \cdot 0}{-3 + 0} = -\infty,$$

$$\lim_{x \rightarrow \infty} \frac{x^3 - 2x^2 + 1}{-x^3 + 1} = \lim_{x \rightarrow \infty} \frac{1 - 2\frac{1}{x} + \frac{1}{x^3}}{-1 + \frac{1}{x^3}} = \frac{2 - 2 \cdot 0 + 0^3}{-1 + 0} = -2.$$

In general, we have the limit

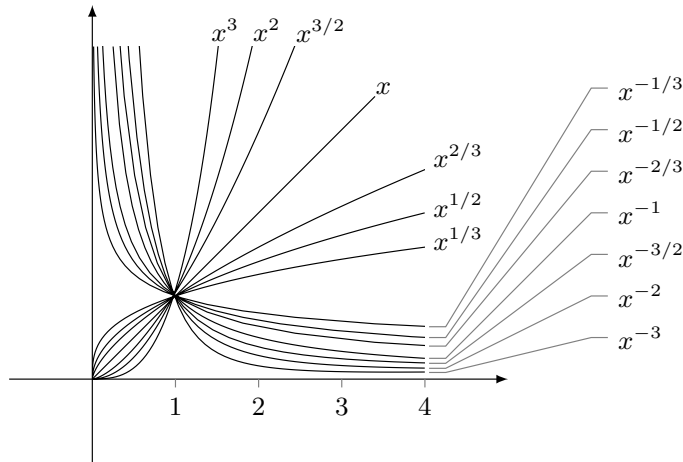
$$\lim_{x \rightarrow \infty} \frac{c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0}{d_m x^m + d_{m-1} x^{m-1} + \cdots + d_1 x + d_0} = \begin{cases} 0, & \text{if } m > n, d_m \neq 0, \\ \frac{c_n}{d_m}, & \text{if } m = n, d_m \neq 0, \\ \infty, & \text{if } m < n, c_n \neq 0. \end{cases}$$

It is also possible to find more refined information on whether the limit is  $+\infty$  or  $-\infty$ . The detail is left to the reader.

## Power Function

The *power function*  $x^p$  is defined for  $x > 0$  and any  $p$  in general. The function can be extended to other cases. For example, the function is defined for all  $x \neq 0$  if  $p$  is an integer. The function is also defined at 0 if  $p \geq 0$ .

If the *exponent*  $p$  is an integer, then the power function is a special case of the rational function, for which we already know the limit. In particular, we have  $\lim_{x \rightarrow a} x^p = a^p$  in case  $p$  is an integer.



**Figure 2.2.1.** Power function.

Now for any  $p$ , fix a natural number  $P > |p|$ . Then for  $x > 1$ , we have  $x^{-P} < x^p < x^P$ . By  $\lim_{x \rightarrow 1} x^{-P} = 1^{-P} = 1$ ,  $\lim_{x \rightarrow 1} x^P = 1^P = 1$ , and the

sandwich rule, we get  $\lim_{x \rightarrow 1^+} x^p = 1$ . Note that the limit is taken from the right because the sandwich inequality holds for  $x > 1$  only. Similarly, by the inequality  $x^{-p} > x^p > x^p$  for  $0 < x < 1$  and the sandwich rule, we get  $\lim_{x \rightarrow 1^-} x^p = 1$ . Combining the limits on two sides, we get  $\lim_{x \rightarrow 1} x^p = 1$ .

For the limit of the power function at any  $a > 0$ , we move the location of the limit from  $a$  to 1

$$\lim_{x \rightarrow a} x^p = \lim_{x \rightarrow 1} (ax)^p = a^p \lim_{x \rightarrow 1} x^p = a^p. \quad (2.2.1)$$

In the first equality, the composition rule is used for

$$x \mapsto y = ax \mapsto z = y^p = (ax)^p.$$

We have  $\lim_{x \rightarrow 1} ax = a$  and  $x \neq 1 \iff y \neq a$ , so that the first additional condition is satisfied. In the second equality, we used the compatibility property  $(ay)^p = a^p y^p$  between the multiplication and the exponential, and the arithmetic rule for the limit.

The limit (2.2.1) shows that, like rational functions, the limit of a power function is the value of the function.

Now we discuss the limits at  $0^+$  and  $+\infty$ . If  $p > 0$ , then for any  $\epsilon > 0$ , we have

$$0 < x < \delta = \epsilon^{\frac{1}{p}} \implies 0 < x^p < \delta^p = \epsilon.$$

This shows that  $\lim_{x \rightarrow 0^+} x^p = 0$ , which can be considered as an extension of Example 1.1.1. The case  $p < 0$  can then be obtained by the arithmetic rule

$$\lim_{x \rightarrow 0^+} x^p = \lim_{x \rightarrow 0^+} \frac{1}{x^{-p}} = \frac{1}{\lim_{x \rightarrow 0^+} x^{-p}} = \frac{1}{0^+} = +\infty \text{ for } p < 0.$$

Combining with  $x^0 = 1$  for all  $x$ , we get

$$\lim_{x \rightarrow 0^+} x^p = \begin{cases} 0, & \text{if } p > 0, \\ 1, & \text{if } p = 0, \\ +\infty, & \text{if } p < 0. \end{cases}$$

By substituting the variable  $x$  with  $\frac{1}{x}$ , we get

$$\lim_{x \rightarrow +\infty} x^p = \begin{cases} 0, & \text{if } p < 0, \\ 1, & \text{if } p = 0, \\ +\infty, & \text{if } p > 0. \end{cases}$$

Note that a substitution of variable is exactly the composition. Therefore computing the limit by substitution makes use of the composition rule. For example, the substituting above means introducing the composition

$$x \mapsto y = \frac{1}{x} \mapsto z = y^{-p} = \frac{1}{x^{-p}} = x^p.$$

We already know the limit  $\lim_{y \rightarrow 0^+} y^{-p}$  of the second function. For the first function, we know  $x \rightarrow +\infty \implies y \rightarrow 0, y > 0$ . A modified version of the composition rule can then be applied.

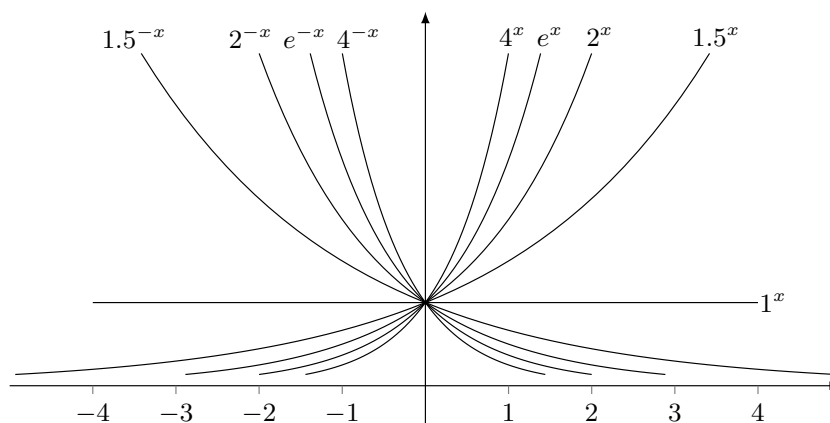


Exercise 2.15. Prove that if  $\lim_{n \rightarrow \infty} x_n = l > 0$ , then  $\lim_{n \rightarrow \infty} x_n^p = l^p$ . What if  $l = 0^+$ ? What if  $l = +\infty$ ?

Exercise 2.16. Prove that if  $\lim_{x \rightarrow a} f(x) = l > 0$ , then  $\lim_{x \rightarrow a} f(x)^p = l^p$ . What if  $l = 0^+$ ? What if  $l = +\infty$ ?

## Exponential Function

The *exponential function*  $c^x$  is defined for  $c > 0$  and all  $x$ .



**Figure 2.2.2.** *Exponential function.*

Example 1.2.3 tells us  $\lim_{n \rightarrow \infty} c^{\frac{1}{n}} = 1$ . This suggests that  $\lim_{x \rightarrow 0^+} c^x = 1$ . For the case  $c \geq 1$ , we prove this by comparing with the known sequence limit.

For  $0 < x < 1$ , we have  $x < \frac{1}{n}$  for some natural number  $n$ . If  $c \geq 1$ , then

$$1 \leq c^x \leq c^{\frac{1}{n}}.$$

Since  $c^{\frac{1}{n}}$  converges to 1, it appears that we can conclude the limit of  $c^x$  by the sandwich rule. However, we cannot directly cite our existing sandwich rules because they do not compare sequences with functions. Instead, we need to repeat the proof of the sandwich rule.

By  $\lim_{n \rightarrow \infty} c^{\frac{1}{n}} = 1$ , for any  $\epsilon > 0$ , there is  $N$ , such that  $n > N$  implies  $|c^{\frac{1}{n}} - 1| < \epsilon$ . Then

$$\begin{aligned} 0 < x < \delta = \frac{1}{N+1} &\implies 0 < x < \frac{1}{n} \text{ for some natural number } n > N \\ &\implies 1 \leq c^x \leq c^{\frac{1}{n}} \text{ (because } c \geq 1) \\ &\implies |c^x - 1| \leq |c^{\frac{1}{n}} - 1| < \epsilon. \end{aligned}$$

This completes the proof of  $\lim_{x \rightarrow 0^+} c^x = 1$  in case  $c \geq 1$ .

For the case  $0 < c \leq 1$ , we have  $\frac{1}{c} \geq 1$ , and the arithmetic rule gives us

$$\lim_{x \rightarrow 0^+} c^x = \frac{1}{\lim_{x \rightarrow 0^+} \left(\frac{1}{c}\right)^x} = 1.$$

Further by substituting  $x$  with  $-x$  (so the composition rule is used), we have

$$\lim_{x \rightarrow 0^-} c^x = \lim_{x \rightarrow 0^+} c^{-x} = \frac{1}{\lim_{x \rightarrow 0^+} c^x} = 1.$$

This completes the proof that  $\lim_{x \rightarrow 0} c^x = 1$  for all  $c > 0$ .

What about the limit at any  $a$ ? Like (2.2.1), we may move the location of the limit from  $a$  to 0, where we already know the limit

$$\lim_{x \rightarrow a} c^x = \lim_{x \rightarrow 0} c^{a+x} = c^a \lim_{x \rightarrow 0} c^x = c^a.$$

Again, the limit of the exponential function is the same as the value of the function.

For the limits at  $\pm\infty$ , Example 1.1.3 says  $\lim_{n \rightarrow \infty} c^n = 0$  for  $|c| < 1$ , which suggests  $\lim_{x \rightarrow +\infty} c^x = 0$  for  $0 < c < 1$ . Again this can be established by the spirit of the sandwich rule. For any  $\epsilon > 0$ , by  $\lim_{n \rightarrow \infty} c^n = 0$ , there is  $N$ , such that  $n > N$  implies  $|c^n| < \epsilon$ . Then

$$\begin{aligned} x > N + 1 &\implies x > n \text{ for some natural number } n > N \\ &\implies 0 < c^x < c^n \\ &\implies |c^x| < |c^n| < \epsilon. \end{aligned}$$

By further applying the arithmetic rule and substitution of variable (composition rule used), we get

$$\lim_{x \rightarrow +\infty} c^x = \begin{cases} 0, & \text{if } 0 < c < 1, \\ 1, & \text{if } c = 1, \\ +\infty, & \text{if } c > 1, \end{cases} \quad (2.2.2)$$

and

$$\lim_{x \rightarrow -\infty} c^x = \begin{cases} 0, & \text{if } c > 1, \\ 1, & \text{if } c = 1, \\ +\infty, & \text{if } 0 < c < 1. \end{cases} \quad (2.2.3)$$

The details are left to readers.

**Exercise 2.17.** Use the limit (1.1.5) to prove

$$\lim_{x \rightarrow +\infty} x^p c^x = 0 \text{ for } 0 < c < 1.$$

Then discuss all cases for  $\lim_{x \rightarrow +\infty} x^p c^x$ .

Exercise 2.18. Use Example 1.1.2 to prove

$$\lim_{x \rightarrow 0^+} x^x = 1,$$

Then prove that  $\lim_{x \rightarrow 0^+} |p(x)|^x = 1$  for any nonzero polynomial  $p(x)$ .

Exercise 2.19. For any  $c > 0$ , we have  $x < c < \frac{1}{x}$  for sufficiently small  $x > 0$ . Use this to derive  $\lim_{x \rightarrow 0} c^x = 1$  from Exercise 2.18.

Exercise 2.20. Prove that if  $\lim_{n \rightarrow \infty} x_n = l$  and  $c > 0$ , then  $\lim_{n \rightarrow \infty} c^{x_n} = c^l$ . What if  $l = +\infty$ ? What if  $l = -\infty$ ?

Exercise 2.21. Prove that if  $\lim_{x \rightarrow a} f(x) = l$  and  $c > 0$ , then  $\lim_{x \rightarrow a} c^{f(x)} = c^l$ .

Exercise 2.22. Prove that if  $0 < A \leq f(x) \leq B$  and  $\lim_{x \rightarrow a} g(x) = 0$ , then  $\lim_{x \rightarrow a} f(x)^{g(x)} = 1$ .

Exercise 2.23. Prove the *exponential rule*: If  $\lim_{x \rightarrow a} f(x) = l > 0$  and  $\lim_{x \rightarrow a} g(x) = k$ , then  $\lim_{x \rightarrow a} f(x)^{g(x)} = l^k$ .

Exercise 2.24. State and prove the exponential rule for sequences.

## Natural Constant $e$

Another basic limit is the extension of the natural constant in Example 1.4.4

$$\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x = e. \quad (2.2.4)$$

Logically, we must derive the limit from the definition (1.4.1), which is the only thing we currently know about  $e$ .

For  $x > 1$ , we have  $n \leq x \leq n + 1$  for some natural number  $n$ . This implies

$$\left(1 + \frac{1}{n+1}\right)^n \leq \left(1 + \frac{1}{x}\right)^x \leq \left(1 + \frac{1}{n}\right)^{n+1}. \quad (2.2.5)$$

From the definition of  $e$ , we know

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^{n+1} &= \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) = e, \\ \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n+1}\right)^n &= \frac{\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n+1}\right)^{n+1}}{\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n+1}\right)} = e. \end{aligned}$$

Therefore for any  $\epsilon > 0$ , there is  $N > 0$ , such that

$$n > N \implies \left| \left(1 + \frac{1}{n}\right)^{n+1} - e \right| < \epsilon, \quad \left| \left(1 + \frac{1}{n+1}\right)^n - e \right| < \epsilon, \quad (2.2.6)$$

Then for  $x > N + 1$ , we have  $n \leq x \leq n + 1$  for some natural number  $n > N$ . By the inequalities (2.2.5) and (2.2.6), this further implies

$$-\epsilon < \left(1 + \frac{1}{n+1}\right)^n - e \leq \left(1 + \frac{1}{x}\right)^x - e \leq \left(1 + \frac{1}{n}\right)^{n+1} - e < \epsilon.$$

This proves  $\lim_{x \rightarrow +\infty} \left(1 + \frac{1}{x}\right)^x = e$ . By substituting  $x$  with  $-x$ , we further get

$$\begin{aligned} \lim_{x \rightarrow -\infty} \left(1 + \frac{1}{x}\right)^x &= \lim_{x \rightarrow +\infty} \left(1 - \frac{1}{x}\right)^{-x} \\ &= \lim_{x \rightarrow +\infty} \left(1 + \frac{1}{x-1}\right)^{x-1} \left(1 + \frac{1}{x-1}\right) = e. \end{aligned}$$

This completes the proof of (2.2.4). Note that substituting  $x$  with  $\frac{1}{x}$  gives us

$$\lim_{x \rightarrow 0} (1+x)^{\frac{1}{x}} = e. \quad (2.2.7)$$

## Trigonometric Function

The sine and tangent functions are defined in Figure 2.2.3, at least for  $0 \leq x \leq \frac{\pi}{2}$ . We have

$$\text{Area}(\text{triangle } OBP) < \text{Area}(\text{fan } OBP) < \text{Area}(\text{triangle } OBQ).$$

This means

$$\frac{1}{2} \sin x < \frac{1}{2} x < \frac{1}{2} \tan x \text{ for } 0 < x < \frac{\pi}{2}.$$

This gives us

$$0 < \sin x < x \text{ for } 0 < x < \frac{\pi}{2}, \quad (2.2.8)$$

and

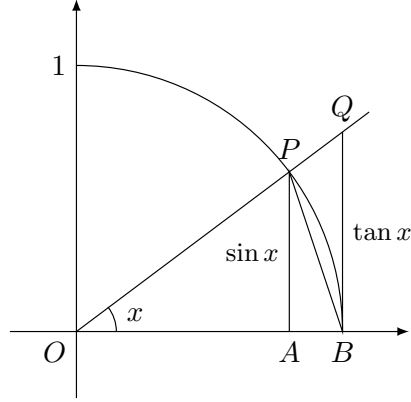
$$\cos x < \frac{\sin x}{x} < 1 \text{ for } 0 < x < \frac{\pi}{2}. \quad (2.2.9)$$

Applying the sandwich rule to (2.2.8), we get  $\lim_{x \rightarrow 0^+} \sin x = 0$ . By substituting  $x$  with  $-x$ , we get  $\lim_{x \rightarrow 0^-} \sin x = \lim_{-x \rightarrow 0^-} \sin(-x) = -\lim_{x \rightarrow 0^+} \sin x = 0$ . This proves

$$\lim_{x \rightarrow 0} \sin x = 0.$$

By the arithmetic rule and substituting  $x$  with  $\frac{x}{2}$ , we get

$$\lim_{x \rightarrow 0} \cos x = \lim_{x \rightarrow 0} \left(1 - 2 \sin^2 \frac{x}{2}\right) = 1 - 2 \left(\lim_{x \rightarrow 0} \sin \frac{x}{2}\right)^2 = 1.$$

**Figure 2.2.3.** *trigonometric function*

Then by trigonometric identities, we have

$$\begin{aligned}
 \lim_{x \rightarrow a} \sin x &= \lim_{x \rightarrow 0} \sin(a + x) = \lim_{x \rightarrow 0} (\sin a \cos x + \cos a \sin x) \\
 &= \sin a \cdot 1 + \cos a \cdot 0 = \sin a, \\
 \lim_{x \rightarrow a} \cos x &= \lim_{x \rightarrow 0} \cos(a + x) = \lim_{x \rightarrow 0} (\cos a \cos x - \sin a \sin x) \\
 &= \cos a \cdot 1 - \sin a \cdot 0 = \cos a, \\
 \lim_{x \rightarrow a} \tan x &= \frac{\lim_{x \rightarrow a} \sin x}{\lim_{x \rightarrow a} \cos x} = \frac{\sin a}{\cos a} = \tan a.
 \end{aligned}$$

The limits of trigonometric functions are always the values of the functions.

Applying  $\lim_{x \rightarrow 0} \cos x = 1$  and the sandwich rule to (2.2.9), we get

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1. \quad (2.2.10)$$

This further implies

$$\lim_{x \rightarrow 0} \frac{\tan x}{x} = \frac{\lim_{x \rightarrow 0} \frac{\sin x}{x}}{\lim_{x \rightarrow 0} \cos x} = 1, \quad (2.2.11)$$

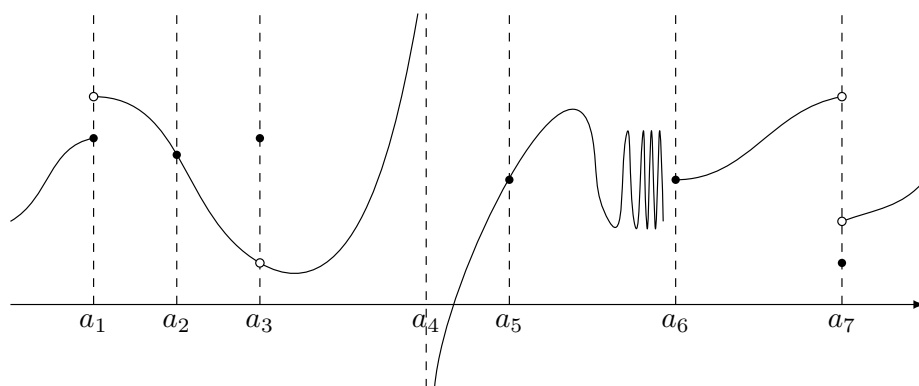
$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \lim_{x \rightarrow 0} \frac{2 \sin^2 \frac{x}{2}}{x^2} = \lim_{x \rightarrow 0} \frac{2 \sin^2 x}{(2x)^2} = \frac{1}{2} \left( \lim_{x \rightarrow 0} \frac{\sin x}{x} \right)^2 = \frac{1}{2}. \quad (2.2.12)$$

## 2.3 Continuity

Changing quantities are often described by functions. Most changes in the real world are smooth, gradual and well behaved. For example, people do not often press the brake when driving a car, and the climate does not suddenly change from summer to winter. The functions describing such well behaved changes are at least continuous.

A function is continuous if its graph “does not break”. The graph of a function  $f(x)$  may break at  $a$  for various reasons. But the breaks are always one of the two types: Either  $\lim_{x \rightarrow a} f(x)$  diverges or the limit converges but the limit value is not  $f(a)$ .

In Figure 2.3.1, the function is continuous at  $a_2, a_5$  and is not continuous at the other five points. Specifically,  $\lim_{x \rightarrow a_1} f(x)$  and  $\lim_{x \rightarrow a_7} f(x)$  diverge because the left and right limits are not equal,  $\lim_{x \rightarrow a_3} f(x)$  converges but not to  $f(a_1)$ ,  $\lim_{x \rightarrow a_4} f(x)$  diverges because the function is not bounded near  $a_4$ , and  $\lim_{x \rightarrow a_6} f(x)$  diverges because the left limit diverges.



**Figure 2.3.1.** *Continuity and discontinuity.*

**Definition 2.3.1.** A function  $f(x)$  defined near and include  $a$  is *continuous* at  $a$  if  $\lim_{x \rightarrow a} f(x) = f(a)$ .

Using the  $\epsilon$ - $\delta$  language, the continuity of  $f(x)$  at  $a$  means that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|x - a| < \delta \implies |f(x) - f(a)| < \epsilon. \quad (2.3.1)$$

A function is *right continuous* at  $a$  if  $\lim_{x \rightarrow a^+} f(x) = f(a)$ , and is *left continuous* if  $\lim_{x \rightarrow a^-} f(x) = f(a)$ . For example, the function in Figure 2.3.1 is left continuous at  $a_1$  and right continuous at  $a_6$ , although it is not continuous at the two points. A function is continuous at  $a$  if and only if it is both left and right continuous at  $a$ .

A function defined on an open interval  $(a, b)$  is continuous if it is continuous at every point on the interval. A function defined on a closed interval  $[a, b]$  is continuous if it is continuous at every point on  $(a, b)$ , is right continuous at  $a$ , and is left continuous at  $b$ . Continuity for functions on other kinds of intervals can be similarly defined.

Most basic functions are continuous. Section 2.2 shows that polynomials, rational functions, trigonometric functions, power functions, and exponential functions are continuous (at the places where the functions are defined). Then the arithmetic rule and the composition rule further imply the following.

**Proposition 2.3.2.** *The arithmetic combinations and the compositions of continuous functions are still continuous.*

So functions such as  $\sin^2 x + \tan x^2$ ,  $\frac{x}{\sqrt{x+1}}$ ,  $2^x \cos \frac{\tan x}{x^2+1}$  are continuous. Examples of discontinuity can only be found among more exotic functions.

Proposition 2.1.7 implies the following criterion for the continuity in terms of sequences.

**Proposition 2.3.3.** *A function  $f(x)$  is continuous at  $a$  if and only if  $\lim_{n \rightarrow \infty} f(x_n) = f(a)$  for any sequence  $x_n$  converging to  $a$ .*

The conclusion of the proposition can be written as

$$\lim_{n \rightarrow \infty} f(x_n) = f\left(\lim_{n \rightarrow \infty} x_n\right). \quad (2.3.2)$$

On the other hand, the condition  $g(b) = c$  in the composition rule in Proposition 2.1.6 simply means that  $g(y)$  is continuous at  $b$ . If we exchange the notations  $f$  and  $g$ , then the composition rule says that the continuity of  $f$  implies

$$\lim_{y \rightarrow b} f(g(y)) = c = f(b) = f\left(\lim_{y \rightarrow b} g(y)\right). \quad (2.3.3)$$

Therefore the function and the limit can be interchanged *if the function is continuous*.

**Example 2.3.1.** The sign function

$$\text{sign}(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -1, & \text{if } x < 0, \end{cases}$$

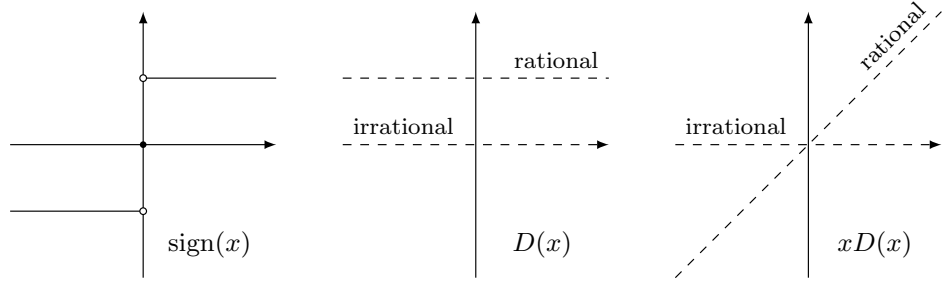
is continuous everywhere except at 0. The Dirichlet function in Example 2.1.2 is not continuous everywhere. Multiplying  $x$  to the Dirichlet function produces a function

$$xD(x) = \begin{cases} x, & \text{if } x \text{ is rational,} \\ 0, & \text{if } x \text{ is irrational,} \end{cases}$$

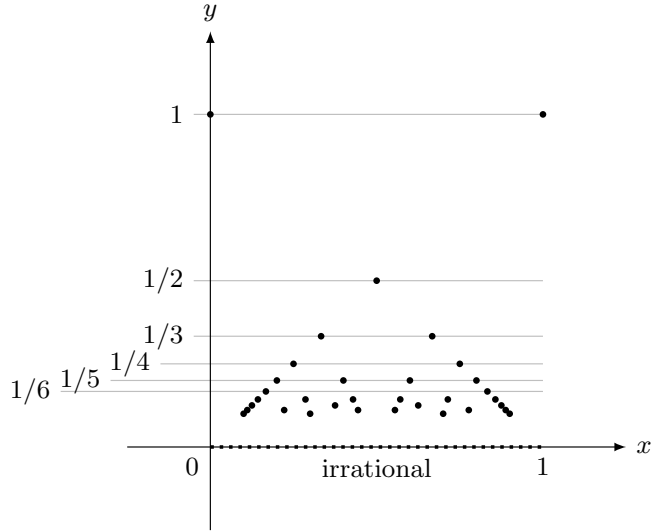
that is continuous only at 0.

**Example 2.3.2** (Thomae<sup>9</sup>). For a rational number  $x = \frac{p}{q}$ , where  $p$  is an integer,  $q$  is a natural number, and  $p, q$  are coprime, define  $R(x) = \frac{1}{q}$ . For an irrational number  $x$ , define  $R(x) = 0$ . Finally define  $R(0) = 1$ . We will show that  $R(x)$  is continuous at precisely all the irrational numbers.

<sup>9</sup>Karl Johannes Thomae, born 1840 in Laucha (Germany), died 1921 in Jena (Germany). Thomae made important contributions to the function theory. In 1870 he showed the continuity in each variable does not imply the joint continuity. He constructed the example here in 1875.



**Figure 2.3.2.** *Discontinuous functions.*



**Figure 2.3.3.** *Thomae's function.*

Let  $a$  be a rational number. Then  $R(a) \neq 0$ . On the other hand, we can find irrational numbers  $x_n$  converging to  $a$ . Then  $\lim_{n \rightarrow \infty} R(x_n) = 0 \neq R(a)$ . By Proposition 2.3.3, the function is not continuous at  $a$ .

Let  $a$  be an irrational number. By the way  $R(x)$  is defined, for any natural number  $N$ , the numbers  $x$  satisfying  $R(x) \geq \frac{1}{N}$  are those rational numbers  $\frac{p}{q}$  with  $q \leq N$ . Therefore on any bounded interval, we have  $R(x) < \frac{1}{N}$  for all except finitely many rational numbers. Let  $x_1, x_2, \dots, x_k$  be all such numbers on the interval  $(a-1, a+1)$ . Then  $x_i \neq a$  because these numbers are rational and  $a$  is irrational. This implies that the smallest distance

$$\delta = \min\{|x_1 - a|, |x_2 - a|, \dots, |x_k - a|, 1\}$$

between  $a$  and these rational numbers is positive. If  $|x - a| < \delta$ , then  $x \in (a-1, a+1)$ , and  $x$  is not equal to any  $x_i$ , which means that  $|R(x) - R(a)| = R(x) < \frac{1}{N}$ . This proves that  $f(x)$  is continuous at  $a$ .



Exercise 2.25. Construct functions on  $(0, 2)$  satisfying the requirements.

1.  $f(x)$  is not continuous at  $\frac{1}{2}$ , 1 and  $\frac{3}{2}$  and is continuous everywhere else.
2.  $f(x)$  is continuous at  $\frac{1}{2}$ , 1 and  $\frac{3}{2}$  and is not continuous everywhere else.
3.  $f(x)$  is continuous everywhere except at  $\frac{1}{n}$  for all natural numbers  $n$ .
4.  $f(x)$  is not left continuous at  $\frac{1}{2}$ , not right continuous at 1, neither side continuous at  $\frac{3}{2}$ , and continuous everywhere else (including the right of  $\frac{1}{2}$  and the left of 1).

Exercise 2.26. Prove that a function  $f(x)$  is continuous at  $a$  if and only if there is  $l$ , such that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|x - a| < \delta \implies |f(x) - l| < \epsilon.$$

By (2.3.1), all you need to do here is to show  $l = f(a)$ .

Exercise 2.27. Prove that a function  $f(x)$  is continuous at  $a$  if and only if for any  $\epsilon > 0$ , there is an interval  $(b, c)$  containing  $a$ , such that

$$x \in (b, c) \implies |f(x) - f(a)| < \epsilon.$$

Exercise 2.28 (Cauchy Criterion). Prove that a function  $f(x)$  is continuous at  $a$  if and only if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|x - a| < \delta, |y - a| < \delta \implies |f(x) - f(y)| < \epsilon.$$

Exercise 2.29. Prove that if a function is continuous on  $(a, b]$  and  $[b, c)$ , then it is continuous on  $(a, c)$ . What about other types of intervals?

Exercise 2.30. Suppose  $f(x)$  and  $g(x)$  are continuous. Prove that  $\max\{f(x), g(x)\}$  and  $\min\{f(x), g(x)\}$  are also continuous.

Exercise 2.31. Suppose  $f(x)$  is continuous on  $[a, b]$  and  $f(r) = 0$  for all rational numbers  $r \in [a, b]$ . Prove that  $f(x) = 0$  on the whole interval.

Exercise 2.32. Suppose for any  $\epsilon > 0$ , only finitely many  $x$  satisfies  $|f(x)| \geq \epsilon$ . Prove that  $\lim_{x \rightarrow a} f(x) = 0$  at any  $a$ . In particular,  $f(x)$  is continuous at  $a$  if and only if  $f(a) = 0$ .

Exercise 2.33. Prove that a continuous function  $f(x)$  on  $(a, b)$  is the restriction of a continuous function on  $[a, b]$  if and only if  $\lim_{x \rightarrow a^+} f(x)$  and  $\lim_{x \rightarrow b^-} f(x)$  converge.

Exercise 2.34. Suppose  $f(x)$  and  $g(x)$  are continuous functions on  $(a, b)$ . Find the places where the following function is continuous

$$h(x) = \begin{cases} f(x), & \text{if } x \text{ is rational,} \\ g(x), & \text{if } x \text{ is irrational.} \end{cases}$$

**Exercise 2.35.** Suppose  $f(x)$  has left limit at every point. Prove that  $g(x) = f(x^-)$  is left continuous, and  $f(x^-) = g(x^-)$ . Similarly, if  $f(x)$  has right limit at every point, then  $h(x) = f(x^+)$  is right continuous, and  $f(x^+) = h(x^+)$ .

**Exercise 2.36.** Suppose  $f(x)$  has left limit and right limit at every point. Let  $g(x) = f(x^-)$  and  $h(x) = f(x^+)$ . Prove that  $f(x^+) = g(x^+)$  and  $f(x^-) = h(x^-)$ .

**Exercise 2.37.** Suppose  $f(x)$  is an increasing function on  $[a, b]$ . Prove that if any number in  $[f(a), f(b)]$  can be the value of  $f(x)$ , then the function is continuous.

**Exercise 2.38.** Suppose  $f(x)$  is an increasing function on  $[a, b]$ . By Proposition 2.1.8, the limits  $f(c^+) = \lim_{x \rightarrow c^+} f(x)$  and  $f(c^-) = \lim_{x \rightarrow c^-} f(x)$  exist at any  $c \in (a, b)$ .

1. Prove that for any  $\epsilon > 0$ , there are finitely many  $c$  satisfying  $f(c^+) - f(c^-) > \epsilon$ .
2. Prove that  $f(x)$  is not continuous only at countably many points.

## 2.4 Compactness Property

The results in this section are stated for closed and bounded intervals. However, the proofs only make use of the following property: Any sequence in  $X$  has a convergent subsequence, and the limit still lies in  $X$ . A set with such a property is called *compact*, so that all the results of this section are still valid on compact sets.

### Uniform Continuity

**Theorem 2.4.1.** Suppose  $f(x)$  is a continuous function on a bounded closed interval  $[a, b]$ . Then for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that for  $x, y \in [a, b]$ ,

$$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon. \quad (2.4.1)$$

In the  $\epsilon$ - $\delta$  formulation (2.3.1) of the continuity, only one variable  $x$  is allowed to change. This means that, in addition to being dependent on  $\epsilon$ , the choice of  $\delta$  may also be dependent on the location  $a$  of the continuity. The property (2.4.1) says that the choice of  $\delta$  can be the *same* for all the points on the interval, so that it depends on  $\epsilon$  only. Therefore the property (which may be defined on any set, not just closed and bounded intervals) is called the *uniform continuity*, and the theorem basically says a continuous function on a bounded closed interval is uniformly continuous.

*Proof.* Suppose  $f(x)$  is not uniformly continuous. Then there is  $\epsilon > 0$ , such that for any natural number  $n$ , there are  $x_n, y_n \in [a, b]$ , such that

$$|x_n - y_n| < \frac{1}{n}, \quad |f(x_n) - f(y_n)| \geq \epsilon. \quad (2.4.2)$$

Since the sequence  $x_n$  is bounded by  $a$  and  $b$ , by Bolzano-Weierstrass Theorem (Theorem 1.5.1), there is a subsequence  $x_{n_k}$  converging to  $c$ . Then by the first

inequality in (2.4.2), we have

$$|y_{n_k} - c| \leq |x_{n_k} - c| + |x_{n_k} - y_{n_k}| < |x_{n_k} - c| + \frac{1}{n}.$$

This and  $\lim_{k \rightarrow \infty} x_{n_k} = c$  imply  $\lim_{k \rightarrow \infty} y_{n_k} = c$ .

By  $a \leq x_n \leq b$  and the order rule (Proposition 1.2.4), we have  $a \leq c \leq b$ . Therefore  $f(x)$  is continuous at  $c$ . By Proposition 2.3.3, we have

$$\lim_{k \rightarrow \infty} f(x_{n_k}) = \lim_{k \rightarrow \infty} f(y_{n_k}) = f(c).$$

Then by the second inequality in (2.4.2), we have

$$\epsilon \leq \left| \lim_{k \rightarrow \infty} f(x_{n_k}) - \lim_{k \rightarrow \infty} f(y_{n_k}) \right| = |f(c) - f(c)| = 0.$$

The contradiction shows that it was wrong to assume that the function is not uniformly continuous.  $\square$

**Example 2.4.1.** Consider the function  $x^2$  on  $[0, 2]$ . For any  $\epsilon > 0$ , take  $\delta = \frac{\epsilon}{4}$ . Then for  $x, y \in [0, 2]$ , we have

$$|x - y| < \delta \implies |x^2 - y^2| = |x - y||x + y| \leq 4|x - y| < \epsilon.$$

Thus  $x^2$  is uniformly continuous on  $[0, 2]$ .

Now consider the same function on  $[0, \infty)$ . For any  $\delta > 0$ , take  $x = \frac{1}{\delta}$ ,  $y = \frac{1}{\delta} + \frac{\delta}{2}$ . Then  $|x - y| < \delta$ , but  $|x^2 - y^2| = x\delta + \frac{\delta^2}{4} > x\delta = 1$ . Thus (2.4.1) fails for  $\epsilon = 1$ , and  $x^2$  is not uniformly continuous on  $[0, \infty)$ .

**Example 2.4.2.** Consider the function  $\sqrt{x}$  on  $[1, \infty)$ . For any  $\epsilon > 0$ , take  $\delta = \epsilon$ . Then for  $x, y \geq 1$ , we have

$$|x - y| < \delta \implies |\sqrt{x} - \sqrt{y}| = \frac{|x - y|}{|\sqrt{x} + \sqrt{y}|} \leq \frac{|x - y|}{2} < \epsilon.$$

Thus  $\sqrt{x}$  is uniformly continuous on  $[1, \infty)$ .

By Theorem 2.4.1, we also know  $\sqrt{x}$  is uniformly continuous on  $[0, 1]$ . Then by Exercise 2.40,  $\sqrt{x}$  is uniformly continuous on  $[0, \infty)$ .

**Example 2.4.3.** Consider the function  $\frac{1}{x}$  on  $(0, 1]$ . For any  $1 > \delta > 0$ , take  $x = \delta$  and  $y = \frac{\delta}{2}$ . Then  $|x - y| = \frac{\delta}{2} < \delta$ , but  $\left| \frac{1}{x} - \frac{1}{y} \right| = \frac{1}{\delta} > 1$ . Therefore (2.4.1) fails for  $\epsilon = 1$ . The function is not uniformly continuous on  $(0, 1]$ .

One may suspect that  $\frac{1}{x}$  is not uniformly continuous on  $(0, 1]$  because the function is not bounded. Here is a bounded example. Consider  $\sin \frac{1}{x}$  on  $(0, 1]$ . For any  $\delta > 0$ , we can find  $0 < x, y < \delta$ , such that  $\frac{1}{x} = 2m\pi + \frac{\pi}{2}$  and  $\frac{1}{y} = 2n\pi - \frac{\pi}{2}$  for some natural numbers

$m, n$ . Then  $|x - y| < \delta$  and  $\sin \frac{1}{x} - \sin \frac{1}{y} = 2$ . Therefore (2.4.1) fails for  $\epsilon = 2$  and the function is not uniformly continuous.

**Exercise 2.39.** Determine uniform continuity.

- |  |  |
|--|--|
| 1. $x^2$ on $(0, 3)$ .                     | 6. $\sin x$ on $(-\infty, +\infty)$ .                    |
| 2. $\frac{1}{x}$ on $[1, 3]$ .             | 7. $\sin \frac{1}{x}$ on $(0, 1]$ .                      |
| 3. $\frac{1}{x}$ on $(0, 3]$ .             | 8. $x \sin \frac{1}{x}$ on $(0, 1]$ .                    |
| 4. $\sqrt[3]{x}$ on $(-\infty, +\infty)$ . | 9. $x^x$ on $(0, 1]$ .                                   |
| 5. $x^{\frac{3}{2}}$ on $[1, +\infty)$ .   | 10. $\left(1 + \frac{1}{x}\right)^x$ on $(0, +\infty)$ . |

**Exercise 2.40.** Prove that if a function is uniformly continuous on  $(a, b]$  and  $[b, c)$ , then it is uniformly continuous on  $(a, c)$ . What about other types of intervals?

**Exercise 2.41.** Let  $f(x)$  be a continuous function on  $(a, b)$ .

1. Prove that if  $\lim_{x \rightarrow a^+} f(x)$  and  $\lim_{x \rightarrow b^-} f(x)$  converge, then  $f(x)$  is uniformly continuous.
2. Prove that if  $(a, b)$  is bounded and  $f(x)$  is uniformly continuous, then  $\lim_{x \rightarrow a^+} f(x)$  and  $\lim_{x \rightarrow b^-} f(x)$  converge.
3. Use  $\sqrt{x}$  to show that the bounded condition in the second part is necessary.
4. Use the second part to show that  $\sin \frac{1}{x}$  is not uniformly continuous on  $(0, 1)$ .

**Exercise 2.42.** A function  $f(x)$  is called *Lipschitz*<sup>10</sup> if there is a constant  $L$  such that  $|f(x) - f(y)| \leq L|x - y|$  for any  $x$  and  $y$ . Prove that Lipschitz functions are uniformly continuous.

**Exercise 2.43.** A function  $f(x)$  on the whole real line is called *periodic* if there is a constant  $p$  such that  $f(x + p) = f(x)$  for any  $x$ . The number  $p$  is the *period* of the function. Prove that continuous periodic functions are uniformly continuous.

**Exercise 2.44.** Is the sum of uniformly continuous functions uniformly continuous? What about the product, the maximum and the composition of uniformly continuous functions?

**Exercise 2.45.** Suppose  $f(x)$  is a continuous function on  $[a, b]$ . Prove that

$$g(x) = \sup\{f(t) : a \leq t \leq x\}$$

is continuous. Is  $g(x)$  uniformly continuous?

<sup>10</sup>Rudolf Otto Sigismund Lipschitz, born 1832 in Königsberg (Germany, now Kaliningrad, Russia), died 1903 in Bonn (Germany). He made important contributions in number theory, Fourier series, differential equations, and mechanics.

**Exercise 2.46 (Dirichlet).** For any continuous function  $f(x)$  on a bounded closed interval  $[a, b]$  and  $\epsilon > 0$ , inductively define

$$c_0 = a, \quad c_n = \sup\{c: |f(x) - f(c_{n-1})| < \epsilon \text{ on } [c_{n-1}, c]\}.$$

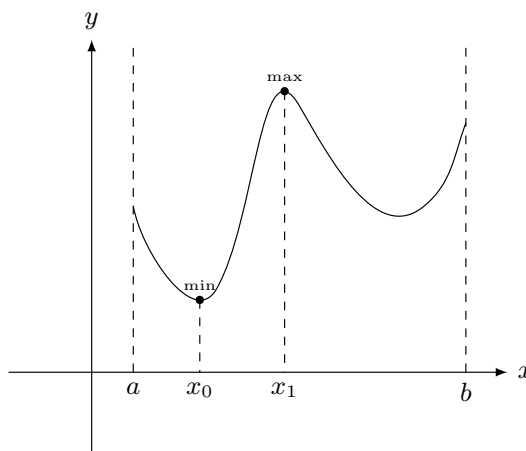
Prove that the process must stop after finitely many steps, which means that there is  $n$ , such that  $|f(x) - f(c_{n-1})| < \epsilon$  on  $[c_n, b]$ . Then use this to give another proof of Theorem 2.4.1.

## Maximum and Minimum

**Theorem 2.4.2.** *A continuous function on a bounded closed interval must be bounded and reaches its maximum and minimum.*

The theorem says that there are  $x_0, x_1 \in [a, b]$ , such that  $f(x_0) \leq f(x) \leq f(x_1)$  for any  $x \in [a, b]$ . The function reaches its minimum at  $x_0$  and its maximum at  $x_1$ .

We also note that, although the theorem tells us the *existence* of the maximum and minimum, it does not tell us how to find them. The maximum and minimum (called *extrema*) can be found by the derivative.



**Figure 2.4.1.** *Maximum and minimum.*

*Proof.* If  $f(x)$  is not bounded on  $[a, b]$ , then there is a sequence  $x_n \in [a, b]$ , such that  $\lim_{n \rightarrow \infty} f(x_n) = \infty$ . By Bolzano-Weierstrass Theorem, there is a convergent subsequence  $x_{n_k}$ . By  $a \leq x_n \leq b$  and the order rule, the limit  $c = \lim_{n \rightarrow \infty} x_{n_k} \in [a, b]$ . Therefore  $f(x)$  is continuous at  $c$ , and  $\lim_{k \rightarrow \infty} f(x_{n_k}) = f(c)$  converges. This contradicts with  $\lim_{n \rightarrow \infty} f(x_n) = \infty$ .

Now the function is bounded. We can introduce  $\beta = \sup\{f(x): x \in [a, b]\}$ . Showing  $f(x)$  reaches its maximum is the same as proving that  $\beta$  is a value of  $f(x)$ . By the characterization of supremum, for any natural number  $n$ , there is  $x_n \in [a, b]$ , such that  $\beta - \frac{1}{n} < f(x_n) \leq \beta$ . By the sandwich rule, we get

$$\lim_{n \rightarrow \infty} f(x_n) = \beta. \quad (2.4.3)$$

On the other hand, since the sequence  $x_n$  is bounded by  $a$  and  $b$ , by Bolzano-Weierstrass Theorem, there is a convergent subsequence  $x_{n_k}$  with  $c = \lim_{k \rightarrow \infty} x_{n_k} \in [a, b]$ . Then we have

$$\lim_{k \rightarrow \infty} f(x_{n_k}) = f(c). \quad (2.4.4)$$

Combining the limits (2.4.3), (2.4.4) and using Proposition 2.3.3, we get  $f(c) = \beta$ . The proof for the function to reach its minimum is similar.  $\square$

**Exercise 2.47.** Construct functions satisfying the requirements.

1.  $f(x)$  is continuous and not bounded on  $(0, 1)$ .
2.  $f(x)$  is continuous and bounded on  $(0, 1)$  but does not reach its maximum.
3.  $f(x)$  is continuous and bounded on  $(0, 1)$ . Moreover,  $f(x)$  also reaches its maximum and minimum.
4.  $f(x)$  is not continuous and not bounded on  $[0, 1]$ .
5.  $f(x)$  is not continuous on  $[0, 1]$  but reaches its maximum and minimum.
6.  $f(x)$  is continuous and bounded on  $(-\infty, \infty)$  but does not reach its maximum.
7.  $f(x)$  is continuous and bounded on  $(-\infty, \infty)$ . Moreover,  $f(x)$  also reaches its maximum and minimum.

What do your examples say about Theorem 2.4.2?

**Exercise 2.48.** Suppose  $f(x)$  is a continuous function on  $(a, b)$ . Prove that if  $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow b^-} f(x) = -\infty$ , then the function reaches its maximum on the interval.

**Exercise 2.49.** Suppose  $f(x)$  is continuous on a bounded closed interval  $[a, b]$ . Suppose for any  $x \in [a, b]$ , there is  $y \in [a, b]$ , such that  $|f(y)| \leq \frac{1}{2}|f(x)|$ . Prove that  $f(c) = 0$  for some  $c \in [a, b]$ . Does the conclusion still hold if the closed interval is changed to an open one?

**Exercise 2.50.** Suppose  $f(x)$  is a uniformly continuous function on a bounded interval  $I$ . Prove that there is  $\delta > 0$ , such that  $f(x)$  is bounded on any interval inside  $I$  of length  $\delta$ . Then prove that  $f(x)$  is bounded on  $I$ .

**Exercise 2.51.** Suppose  $f(x)$  is a uniformly continuous function on  $\mathbb{R}$ . Prove that for any  $a$ ,  $f(x+a) - f(x)$  is bounded.

## 2.5 Connectedness Property

The results of this section are generally valid for all (not just closed and bounded) intervals. The key reason behind the results is that, inside an interval, one can always “continuously move” from one point to another point. A set with such a property is called *connected*.

### Intermediate Value Theorem

**Theorem 2.5.1** (Intermediate Value Theorem). *Suppose  $f(x)$  is a continuous function on a bounded closed interval  $[a, b]$ . If  $y$  is a number between  $f(a)$  and  $f(b)$ , then  $y = f(c)$  for some  $c \in [a, b]$ .*

*Proof.* Without loss of generality, assume  $f(a) \leq y \leq f(b)$ . Let

$$X = \{x \in [a, b] : f(x) \leq y\}.$$

The set is not empty because  $a \in X$ . The set is also bounded by  $a$  and  $b$ . Therefore we have  $c = \sup X \in [a, b]$ . We expect  $f(c) = y$ .

If  $f(c) > y$ , then by the continuity,  $\lim_{x \rightarrow c} f(x) = f(c) > y$ . By the order rule in Proposition 2.1.6, there is  $\delta > 0$ , such that  $f(x) > y$  for any  $x \in (c - \delta, c]$  (the right side of  $c$  may not be allowed in case  $c = b$ ). On the other hand, by  $c = \sup X$ , there is  $x' \in (c - \delta, c]$  satisfying  $f(x') \leq y$ . We get a contradiction at  $x'$ .

If  $f(c) < y$ , then by  $f(b) \geq y$ , we have  $c < b$ . Again by the continuity of  $f(x)$  at  $c$  and the order rule, there is  $\delta > 0$ , such that  $|x - c| < \delta$  implies  $f(x) < y$ . In particular, any  $x' \in (c, c + \delta)$  will satisfy  $x' > c$  and  $f(x') < y$ . This contradicts with the assumption that  $c$  is an upper bound of  $X$ .

Thus we conclude that  $f(c) = y$ , and the proof is complete.  $\square$

For general intervals, we have the following version of the Intermediate Value Theorem.

**Theorem 2.5.2.** *Suppose  $f(x)$  is a continuous function on an interval  $I$ . Then the values  $f(I) = \{f(x) : x \in I\}$  of the function on  $I$  is an interval of left end  $\inf_I f$  and right end  $\sup_I f$ .*

In case  $I = [a, b]$  is a bounded and closed interval, by Theorem 2.4.2, the function reaches its minimum  $\alpha = \inf_{[a, b]} f$  and maximum  $\beta = \sup_{[a, b]} f$ . We conclude that  $f([a, b]) = [\alpha, \beta]$ . In general, whether the interval  $f(I)$  includes the left end  $\alpha$  or the right end  $\beta$  depends on whether the minimum or the maximum is reached.

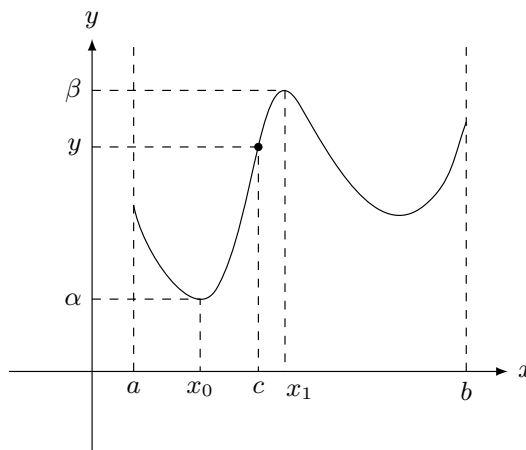
*Proof.* We always have  $f(I) \subset [\alpha, \beta]$  for  $\alpha = \inf_I f$  and maximum  $\beta = \sup_I f$ . We only need to show that any number  $y \in (\alpha, \beta)$  is the value of  $f$ . The assumption  $\inf_I f = \alpha < y < \beta = \sup_I f$  implies that  $f(a) < y < f(b)$  for some  $a, b \in I$ . Applying Theorem 2.5.1 to the continuous function  $f$  on the interval  $[a, b] \subset I$ , we get  $f(c) = y$  for some  $c \in [a, b] \subset I$ .  $\square$

**Example 2.5.1.** For the function  $f(x) = x^5 + 2x^3 - 5x^2 + 1$ , we have  $f(0) = 1$ ,  $f(1) = -1$ ,  $f(2) = 29$ . By the Intermediate Value Theorem, there are  $a \in [0, 1]$  and  $b \in [1, 2]$  such that  $f(a) = f(b) = 0$ .

**Example 2.5.2.** A number  $a$  is a *root* of a function  $f(x)$  if  $f(a) = 0$ . For a polynomial  $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$  with  $a_n > 0$  and  $n$  odd, we have

$$f(x) = x^n g(x), \quad g(x) = a_n + \frac{a_{n-1}}{x} + \cdots + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n}.$$

Since  $\lim_{x \rightarrow \infty} g(x) = a_n > 0$  and  $n$  is odd, we have  $f(x) > 0$  for sufficiently big and positive  $x$ , and  $f(x) < 0$  for sufficiently big and negative  $x$ . Then by the Intermediate Value Theorem,  $f(a) = 0$  for some  $a$  (between two sufficiently big numbers of opposite



**Figure 2.5.1.** *Intermediate value theorem.*

signs). Similar argument also works for the case  $a_n < 0$ . Therefore we conclude that a polynomial of odd degree must have a real root.

**Example 2.5.3.** We claim that  $\lim_{x \rightarrow +\infty} f\left(\frac{\sin x}{x}\right)$  converges if and only if  $f(x)$  is continuous at 0.

Let  $g(x) = \frac{\sin x}{x}$ . If  $f(x)$  is continuous at 0, then by  $\lim_{x \rightarrow +\infty} g(x) = 0$  and the composition rule, we get  $\lim_{x \rightarrow +\infty} f(g(x)) = \lim_{x \rightarrow 0} f(x) = f(0)$ .

Conversely, assume  $\lim_{x \rightarrow +\infty} f(g(x)) = l$ , we need to prove the continuity of  $f$  at 0. The limit  $\lim_{x \rightarrow +\infty} f(g(x)) = l$  means that, for any  $\epsilon > 0$ , there is  $N$ , such that

$$x > N \implies |f(g(x)) - l| < \epsilon.$$

By Exercise 2.26, the continuity of  $f$  at 0 means that, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that (we substituted  $y$  for  $x$ )

$$|y| < \delta \implies |f(y) - l| < \epsilon.$$

The key to proving that the first implication implies the second implication is to express very small  $y$  in the second implication as  $g(x)$  in the first implication.

Suppose for the given  $\epsilon > 0$ , we have  $N$  such that the first implication holds. We can easily find  $x_1, x_2 > N$ , such that  $g(x_1) > 0$  and  $g(x_2) < 0$ . Let for any  $|y| < \delta = \min\{g(x_1), -g(x_2)\}$ , we have  $g(x_1) > y > g(x_2)$ . Applying the Intermediate Value Theorem to the continuous function  $g(x)$  on  $[x_1, x_2]$ , we find  $y = g(x)$  for some  $x \in [x_1, x_2]$ . Since  $x_1, x_2 > N$ , we have  $x > N$ . Then

$$|y| < \delta \implies y = g(x), \text{ for some } x > N \implies |f(y) - l| = |f(g(x)) - l| < \epsilon.$$

Exercises 2.59 and 2.60 extend the example.

**Exercise 2.52.** Prove that for any polynomial of odd degree, any real number is the value of the polynomial.



Exercise 2.53. Show that  $2^x = 3x$  has solution on  $(0, 1)$ . Show that  $3^x = x^2$  has solution.

Exercise 2.54. Suppose a continuous function on an interval is never zero. Prove that it is always positive or always negative.

Exercise 2.55. Suppose  $f(x)$  is a continuous function on  $(a, b)$ . Prove that if  $f(x)$  only takes rational numbers as values, then  $f(x)$  is a constant.

Exercise 2.56. Suppose  $f(x)$  and  $g(x)$  are continuous functions on  $[a, b]$ . Prove that if  $f(a) < g(a)$  and  $f(b) > g(b)$ , then  $f(c) = g(c)$  for some  $c \in (a, b)$ .

Exercise 2.57. Suppose  $f: [0, 1] \rightarrow [0, 1]$  is a continuous function. Prove that  $f(c) = c$  for some  $c \in [0, 1]$ . We call  $c$  a *fixed point* of  $f$ .

Exercise 2.58. Suppose  $f(x)$  is a two-to-one function on  $[a, b]$ . In other words, for any  $x \in [a, b]$ , there is exactly one other  $y \in [a, b]$  such that  $x \neq y$  and  $f(x) = f(y)$ . Prove that  $f(x)$  is not continuous.

Exercise 2.59. Example 2.5.3 basically says that, if  $g(x)$  is a nice continuous function near  $a$ , then the convergence of a composition  $f(g(x))$  as  $x \rightarrow a$  implies the convergence of  $f(x)$  as  $x \rightarrow g(a)$ . Prove another version of this “converse of composition rule”: If  $g(x)$  is continuous near  $a$ , and for any  $\delta > 0$ , we have  $g(x) > g(a)$  for some  $x \in (a - \delta, a + \delta)$ , then

$$\lim_{x \rightarrow a} f(g(x)) = l \implies \lim_{x \rightarrow g(a)^+} f(x) = l.$$

Exercise 2.60. Let  $g(x)$  be a continuous function on  $(a, b)$  with convergent right limit  $\lim_{x \rightarrow a^+} g(x) = \alpha$ . Find suitable condition on  $g(x)$  so that the following implication is true

$$\lim_{x \rightarrow a^+} f(g(x)) = l \implies \lim_{x \rightarrow \alpha^-} f(x) = l.$$

Then show that  $\lim_{x \rightarrow 0^+} f\left(\frac{\sin x}{x}\right) = \lim_{x \rightarrow 1^-} f(x)$  and  $\lim_{x \rightarrow \frac{\pi}{2}} f(\sin x) = \lim_{x \rightarrow 1^-} f(x)$ .

Exercise 2.61. Show that if the function  $g$  in Example 2.5.3 and Exercises 2.59 and 2.60 is a constant, then the conclusion is not true. Moreover, show the continuity of  $g$  is necessary by considering  $g(x) = \frac{1}{n}$  on  $\left(\frac{1}{n+1}, \frac{1}{n}\right]$  and constructing  $f(x)$ , such that  $\lim_{x \rightarrow 0^+} f(g(x))$  converges but  $\lim_{x \rightarrow 0^+} f(x)$  diverges.

## Invertible Continuous Function

Functions are maps. By writing a function in the form  $f: [a, b] \rightarrow [\alpha, \beta]$ , we mean the function is defined on the *domain*  $[a, b]$  and its values lie in the *range*  $[\alpha, \beta]$ . The function is *onto* (or *surjective*) if any  $y \in [\alpha, \beta]$  is the value  $y = f(x)$  at some  $x \in [a, b]$ . It is *one-to-one* (or *injective*) if  $x_1 \neq x_2$  implies  $f(x_1) \neq f(x_2)$ . It is *invertible* (or *bijective*) if there is another function  $g: [\alpha, \beta] \rightarrow [a, b]$  such that  $g(f(x)) = x$  for any  $x \in [a, b]$  and  $f(g(y)) = y$  for any  $y \in [\alpha, \beta]$ . The function  $g$  is called the *inverse* of  $f$  and is denoted  $g = f^{-1}$ . It is a basic fact that a function is

invertible if and only if it is onto and one-to-one. Moreover, the inverse function is unique.

The discussion above also applies to the case the domain and the range are intervals of other kinds. For a continuous function  $f$  on an interval  $I$ , by Theorem 2.5.2, the values  $f(I)$  of the function also form an interval, and the invertibility of  $f$  means the invertibility of the map  $f: I \rightarrow f(I)$  between intervals.

**Theorem 2.5.3.** *A continuous function on an interval is invertible if and only if it is strictly monotone. Moreover, the inverse is also continuous and strictly monotone.*

*Proof.* Since the map  $f: I \rightarrow f(I)$  is always onto,  $f$  is invertible if and only if it is one-to-one.

If  $f(x)$  is strictly increasing, then

$$\begin{aligned} x_1 \neq x_2 &\iff x_1 > x_2 \text{ or } x_1 < x_2 \\ &\implies f(x_1) > f(x_2) \text{ or } f(x_1) < f(x_2) \\ &\iff f(x_1) \neq f(x_2). \end{aligned}$$

This proves that  $f: I \rightarrow f(I)$  is one-to-one and is therefore invertible. By the same reason, strictly decreasing also implies invertible.

Conversely, suppose  $f: I \rightarrow f(I)$  is invertible. Pick any interval  $[a, b]$  in  $I$  and assume  $f(a) \leq f(b)$ . We will prove that  $f(x)$  is strictly increasing on  $[a, b]$ . Assume not. Then there are  $a \leq x_1 < x_2 \leq b$  satisfying  $f(x_1) \geq f(x_2)$ . We consider three possibilities for  $f(x_1)$ .

1. If  $f(x_1) \leq f(a)$ , then  $f(x_2) < f(a) \leq f(b)$ . Applying the Intermediate Value Theorem to the value  $f(a)$  on  $[x_2, b]$ , we get  $f(a) = f(c)$  for some  $c \in [x_2, b]$ . Since  $c \geq x_2 > a$ ,  $f$  has the same value at two distinct places  $a$  and  $c$ .
2. If  $f(a) < f(x_1) \leq f(b)$ , then  $f(x_2) < f(x_1) \leq f(b)$ . Applying the Intermediate Value Theorem to the value  $f(x_1)$  on  $[x_2, b]$ , we get  $f(x_1) = f(c)$  for some  $c \in [x_2, b]$ . Since  $c \geq x_2 > x_1$ ,  $f$  has the same value at two distinct places  $x_1$  and  $c$ .
3. If  $f(b) < f(x_1)$ , then  $f(a) < f(b) < f(x_1)$ . Applying the Intermediate Value Theorem to the value  $f(b)$  on  $[a, x_1]$ , we get  $f(b) = f(c)$  for some  $c \in [a, x_1]$ . Since  $c \leq x_1 < b$ ,  $f$  has the same value at two distinct places  $b$  and  $c$ .

In all cases,  $f(x)$  fails to be one-to-one. The contradiction shows that  $f(x)$  must be strictly increasing on  $[a, b]$ .

Now consider any interval  $[c, d]$  between  $[a, b]$  and  $I$ . Depending on whether  $f(c) \leq f(d)$  or  $f(c) \geq f(d)$ , applying the similar argument to the interval  $[c, d]$  in  $I$  shows that  $f$  is either strictly increasing or strictly decreasing on  $[c, d]$ . Since we already know that  $f$  is strictly increasing on  $[a, b] \subset [c, d]$ ,  $f$  must be strictly increasing on  $[c, d]$ . Since any two points in  $I$  lies in an interval  $[c, d]$  between  $[a, b]$  and  $I$ , we conclude that  $f$  is strictly increasing on  $I$ .

We get strictly increasing on  $I$  under the assumption  $f(a) \leq f(b)$ . By the same reason, we get strictly decreasing on  $I$  under the assumption  $f(a) \geq f(b)$ . This completes the proof that the invertibility implies the strict monotonicity.

It remains to prove that if  $f(x)$  is continuous, strictly increasing and invertible, then its inverse  $f^{-1}(y)$  is also continuous and strictly increasing.

Let  $y_1 = f(x_1)$  and  $y_2 = f(x_2)$ . Then by  $f(x)$  increasing, we have

$$x_1 \geq x_2 \implies y_1 \geq y_2.$$

The implication is the same as

$$y_1 < y_2 \implies x_1 < x_2.$$

By  $x_1 = f^{-1}(y_1)$  and  $x_2 = f^{-1}(y_2)$ , this means exactly that  $f^{-1}(x)$  is strictly increasing. To prove the continuity of  $f^{-1}(x)$  at  $\gamma \in f(I)$ , we apply Proposition 2.1.8 to the strictly increasing function  $f^{-1}$  and get a convergent limit  $\lim_{y \rightarrow \gamma^+} f^{-1}(y) = c$ . By the continuity of  $f(x)$  at  $c$  and (2.3.3), we get

$$\gamma = \lim_{y \rightarrow \gamma^+} y = \lim_{y \rightarrow \gamma^+} f(f^{-1}(y)) = f\left(\lim_{y \rightarrow \gamma^+} f^{-1}(y)\right) = f(c).$$

Therefore  $c = f^{-1}(\gamma)$ , and  $\lim_{y \rightarrow \gamma^+} f^{-1}(y) = c = f^{-1}(\gamma)$ . This means that  $f^{-1}$  is right continuous at  $\gamma$ . By similar reason,  $f^{-1}$  is also left continuous.  $\square$

We remark that if  $f(x)$  is strictly increasing and continuous on  $(a, b)$ , and  $(\alpha, \beta) = f(a, b)$ , then we actually have

$$\lim_{y \rightarrow \alpha^+} f^{-1}(y) = a, \quad \lim_{y \rightarrow \beta^-} f^{-1}(y) = b.$$

The claim is true even if some of  $a, b, \alpha, \beta$  are infinity. Similar remark also applies to the strictly decreasing functions and other types of intervals. See Exercise 2.62.

**Exercise 2.62.** Suppose  $f(x)$  is a strictly decreasing and continuous function on  $[a, b]$ . Let  $\alpha = f(a)$  and  $\beta = \lim_{x \rightarrow b^-} f(x)$ . Prove that  $f: [a, b] \rightarrow (\beta, \alpha]$  is invertible and  $\lim_{y \rightarrow \beta^+} f^{-1}(y) = b$ .

**Exercise 2.63.** Suppose  $g(x)$  is a strictly decreasing and continuous function on  $[a, a + \delta)$ , and  $b = g(a)$ . Prove that  $\lim_{x \rightarrow b^-} f(x) = \lim_{x \rightarrow a^+} f(g(x))$  in the strong sense that, the left side converges if and only if the right side converges, and the two sides are equal when they converge.

## Basic Inverse Functions

The trigonometric functions

$$\sin: \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \rightarrow [-1, 1], \quad \cos: [0, \pi] \rightarrow [-1, 1], \quad \tan: \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \rightarrow (-\infty, +\infty)$$

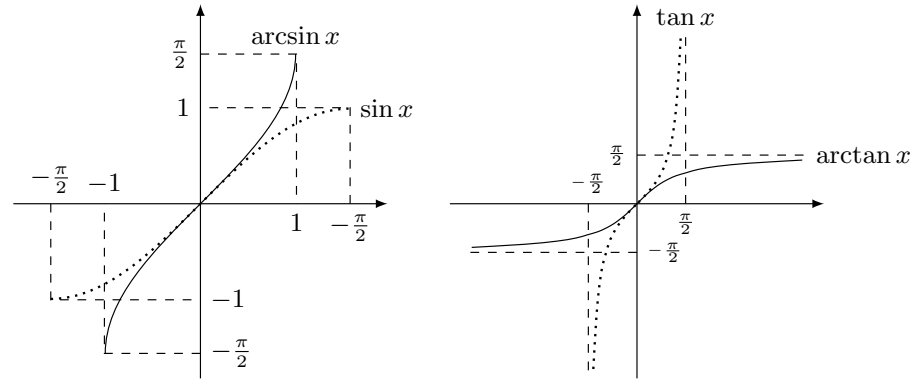
are onto, strictly monotone and continuous. Therefore they are invertible, and the *inverse trigonometric functions*  $\arcsin$ ,  $\arccos$ ,  $\arctan$  are also strictly monotone

and continuous. Although the other inverse trigonometric functions may be defined similarly, the equality  $\cos\left(\frac{\pi}{2} - x\right) = \sin x$  implies that

$$\arcsin x + \arccos x = \frac{\pi}{2}, \quad (2.5.1)$$

and we have similar simple equations relating other inverse trigonometric functions. Moreover, by the remark made after the proof of Theorem 2.5.3, we have

$$\lim_{x \rightarrow -\infty} \arctan x = -\frac{\pi}{2}, \quad \lim_{x \rightarrow +\infty} \arctan x = \frac{\pi}{2}.$$



**Figure 2.5.2.** *Inverse trigonometric functions.*

The exponential function  $a^x$  based on a constant  $a > 0$  is continuous. The function is strictly increasing for  $a > 1$  and strictly decreasing for  $0 < a < 1$ . Moreover, the limits at the infinity are given by (2.2.2) and (2.2.3). Therefore the map

$$a^x: (-\infty, \infty) \rightarrow (0, \infty), \quad 0 < a \neq 1$$

is invertible. The inverse function

$$\log_a x: (0, \infty) \rightarrow (-\infty, \infty), \quad 0 < a \neq 1$$

is the *logarithmic function*, which is also continuous, strictly increasing for  $a > 1$  and strictly decreasing for  $0 < a < 1$ . Moreover, we have

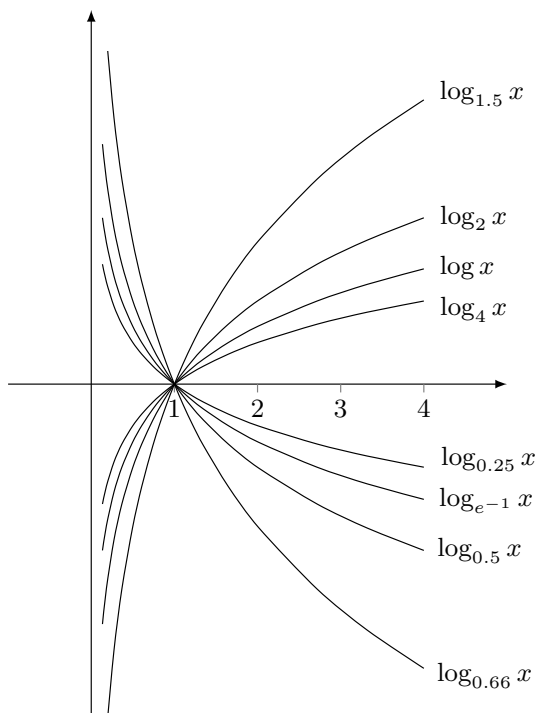
$$\lim_{x \rightarrow 0^+} \log_a x = \begin{cases} -\infty, & \text{if } a > 1, \\ +\infty, & \text{if } 0 < a < 1, \end{cases} \quad \lim_{x \rightarrow +\infty} \log_a x = \begin{cases} +\infty, & \text{if } a > 1, \\ -\infty, & \text{if } 0 < a < 1. \end{cases}$$

The following equalities for the exponential function

$$a^0 = 1, \quad a^1 = a, \quad a^x a^y = a^{x+y}, \quad (a^x)^y = a^{xy},$$

imply the following equalities for the logarithmic function

$$\log_a 1 = 0, \quad \log_a a = 1, \quad \log_a(xy) = \log_a x + \log_a y,$$



**Figure 2.5.3.** *Logarithmic functions.*

$$\log_a x^y = y \log_a x, \quad \log_a x = \frac{\log x}{\log a}.$$

The logarithmic function in the special base  $a = e$  is called the *natural logarithmic function* and is denoted by  $\log$  or  $\ln$ . In particular, we have  $\log e = 1$ . By applying the continuity of  $\log$  to the limit (2.2.7), we get

$$\lim_{x \rightarrow 0} \frac{\log(1+x)}{x} = \lim_{x \rightarrow 0} \log(1+x)^{\frac{1}{x}} = \log \left( \lim_{x \rightarrow 0} (1+x)^{\frac{1}{x}} \right) = \log e = 1. \quad (2.5.2)$$

The continuity of  $e^x$  tells us  $\lim_{x \rightarrow 0} (e^x - 1) = 0$ . Substituting  $x$  in (2.5.2) with  $e^x - 1$ , we get

$$\lim_{x \rightarrow 0} \frac{x}{e^x - 1} = \lim_{x \rightarrow 0} \frac{\log(1 + (e^x - 1))}{e^x - 1} = 1.$$

Taking reciprocal, we get

$$\lim_{x \rightarrow 0} \frac{e^x - 1}{x} = 1. \quad (2.5.3)$$

Substituting  $x$  with  $x \log a$  in (2.5.3) and using  $e^{x \log a} = a^x$ , we get a more general formula

$$\lim_{x \rightarrow 0} \frac{a^x - 1}{x} = \log a. \quad (2.5.4)$$

The continuity of  $\log$  tells us  $\lim_{x \rightarrow 0} \log(1+x) = 0$ . Substituting  $x$  in (2.5.3) with

$p \log(1+x)$  and using  $e^{p \log(1+x)} = (1+x)^p$ , we get

$$\lim_{x \rightarrow 0} \frac{(1+x)^p - 1}{\log(1+x)} = p.$$

Multiplying the limit with (2.5.2), we get

$$\lim_{x \rightarrow 0} \frac{(1+x)^p - 1}{x} = p. \quad (2.5.5)$$

**Exercise 2.64.** Prove  $\lim_{x \rightarrow 1^-} f(\arcsin x) = \lim_{x \rightarrow \frac{\pi}{2}^-} f(x)$  and  $\lim_{x \rightarrow 0} f\left(\frac{\sin x}{\arcsin x}\right) = \lim_{x \rightarrow 1^-} f(x)$ .

**Exercise 2.65.** Use Exercise 1.66 and the continuity of logarithmic function to prove that if  $x_n > 0$  and  $\lim_{n \rightarrow \infty} x_n = l$ , then  $\lim_{n \rightarrow \infty} \sqrt[n]{x_1 x_2 \cdots x_n} = l$ . What about the case  $\lim_{n \rightarrow \infty} x_n = +\infty$ ?

**Exercise 2.66.** Use Exercise 2.17 to derive

$$\lim_{x \rightarrow +\infty} \frac{(\log x)^p}{x} = 0, \quad \lim_{x \rightarrow 0^+} x |\log x|^p = 0.$$

This means that when  $x$  approaches  $+\infty$  or  $0^+$ ,  $\log x$  approaches  $\infty$  but at a much slower speed than the speed of  $x$  approaching its target. Moreover, discuss  $\lim_{x \rightarrow +\infty} a^x x^p (\log x)^q$  and  $\lim_{x \rightarrow 0^+} x^p (\log x)^q$ .

## 2.6 Additional Exercise

### Extended Exponential Rule

**Exercise 2.67.** Prove the extended exponential rules.

1.  $l^{+\infty} = +\infty$  for  $l > 1$ : If  $\lim_{x \rightarrow a} f(x) = l > 1$  and  $\lim_{x \rightarrow a} g(x) = +\infty$ , then  $\lim_{x \rightarrow a} f(x)^{g(x)} = +\infty$ .
2.  $(0^+)^k = 0$  for  $k > 0$ : If  $f(x) > 0$ ,  $\lim_{x \rightarrow a} f(x) = 0$  and  $\lim_{x \rightarrow a} g(x) = k > 0$ , then  $\lim_{x \rightarrow a} f(x)^{g(x)} = 0$ .

From the two rules, further derive the following exponential rules.

1.  $l^{+\infty} = 0$  for  $0 < l < 1$ .
2.  $l^{-\infty} = 0$  for  $l > 1$ .
3.  $(0^+)^k = +\infty$  for  $k < 0$ .
4.  $(+\infty)^k = 0$  for  $k > 0$ .

**Exercise 2.68.** Provide counterexamples to the wrong exponential rules

$$(+\infty)^0 = 1, \quad 1^{+\infty} = 1, \quad 0^0 = 1, \quad 0^0 = 0.$$

### Comparison of Small Sums

If  $f(x)$  and  $g(x)$  are equivalent infinitesimals at 0, then we expect  $f(x_1) + f(x_2) + \cdots + f(x_n)$  and  $g(x_1) + g(x_2) + \cdots + g(x_n)$  to be very close to each other when  $x_1, x_2, \dots, x_n$  are very small. The following exercises indicate some cases our expectation is fulfilled.

Exercise 2.69. Suppose  $f(x)$  is a function on  $(0, 1]$  satisfying  $\lim_{x \rightarrow 0^+} \frac{f(x)}{x} = 1$ . Prove that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left( f\left(\frac{1}{n^2}\right) + f\left(\frac{2}{n^2}\right) + f\left(\frac{3}{n^2}\right) + \cdots + f\left(\frac{n}{n^2}\right) \right) \\ &= \lim_{n \rightarrow \infty} \left( \frac{1}{n^2} + \frac{2}{n^2} + \frac{3}{n^2} + \cdots + \frac{n}{n^2} \right) = \lim_{n \rightarrow \infty} \frac{n+1}{2n} = \frac{1}{2}. \end{aligned}$$

Exercise 2.70. Suppose  $g(x) > 0$  and  $\lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = 1$ . Suppose for each natural number  $n$ , there are nonzero numbers  $x_{n,1}, x_{n,2}, \dots, x_{n,k_n}$ , so that  $\lim_{n \rightarrow \infty} x_{n,k} = 0$  uniformly in  $k$ : For any  $\epsilon > 0$ , there is  $N$ , such that  $n > N$  implies  $|x_{n,k}| < \epsilon$ . Prove that if

$$\lim_{n \rightarrow \infty} (g(x_{n,1}) + g(x_{n,2}) + \cdots + g(x_{n,k_n})) = l,$$

then

$$\lim_{n \rightarrow \infty} (f(x_{n,1}) + f(x_{n,2}) + \cdots + f(x_{n,k_n})) = l.$$

### Upper and Lower Limits of Functions

Suppose  $f(x)$  is defined near (but not necessarily at)  $a$ . Let

$$\text{LIM}_a f = \left\{ \lim_{n \rightarrow \infty} f(x_n) : x_n \neq a, \lim_{n \rightarrow \infty} x_n = a, \lim_{n \rightarrow \infty} f(x_n) \text{ converges} \right\}.$$

Define

$$\overline{\lim}_{x \rightarrow a} f(x) = \sup \text{LIM}_a f, \quad \underline{\lim}_{x \rightarrow a} f(x) = \inf \text{LIM}_a f.$$

Similar definitions can be made when  $a$  is replaced by  $a^+, a^-, \infty, +\infty$  and  $-\infty$ .

Exercise 2.71. Prove the analogue of Proposition 1.5.3:  $l \in \text{LIM}_a f$  if and only if for any  $\epsilon > 0$  and  $\delta > 0$ , there is  $x$  satisfying  $0 < |x - a| < \delta$  and  $|f(x) - l| < \epsilon$ .

Exercise 2.72. Prove the analogue of Proposition 1.5.4: The upper limit is completely characterized by the following two properties.

1. If  $l > \overline{\lim}_{x \rightarrow a} f(x)$ , then there is  $\delta > 0$ , such that  $0 < |x - a| < \delta$  implies  $f(x) \leq l$ .
2. If  $l < \overline{\lim}_{x \rightarrow a} f(x)$ , then for any  $\delta > 0$ , there is  $x$  satisfying  $0 < |x - a| < \delta$  and  $f(x) > l$ .

Note that for the function limit, the properties are not equivalent to the existence of finitely or infinitely many  $x$ .

Exercise 2.73. Prove the analogue of Proposition 1.5.5: The upper and lower limits of  $f(x)$  at  $a$  belong to  $\text{LIM}_a f$ , and  $\lim_{x \rightarrow a} f(x)$  converges if and only if the upper and lower limits are equal.

Exercise 2.74. Prove the analogue of Exercise 1.54:

$$\begin{aligned} \overline{\lim}_{x \rightarrow a} f(x) &= \lim_{\delta \rightarrow 0^+} \sup \{f(x) : 0 < |x - a| < \delta\}, \\ \underline{\lim}_{x \rightarrow a} f(x) &= \lim_{\delta \rightarrow 0^+} \inf \{f(x) : 0 < |x - a| < \delta\}. \end{aligned}$$

**Exercise 2.75.** Prove the analogue of Exercise 1.47: If  $l_k \in \text{LIM}_a f$  and  $\lim_{k \rightarrow \infty} l_k = l$ , then  $l \in \text{LIM}_a f$ .

**Exercise 2.76.** Prove the extension of Proposition 2.1.3:

$$\text{LIM}_a f = \text{LIM}_{a^+} f \cup \text{LIM}_{a^-} f.$$

In particular, we have

$$\begin{aligned} \overline{\lim}_{x \rightarrow a} f(x) &= \max \left\{ \overline{\lim}_{x \rightarrow a^+} f(x), \overline{\lim}_{x \rightarrow a^-} f(x) \right\}, \\ \underline{\lim}_{x \rightarrow a} f(x) &= \min \left\{ \underline{\lim}_{x \rightarrow a^+} f(x), \underline{\lim}_{x \rightarrow a^-} f(x) \right\}. \end{aligned}$$

**Exercise 2.77.** Extend the arithmetic and order properties of upper and lower limits in Exercise 1.49.

### Additive and Multiplicative Functions

**Exercise 2.78.** Suppose  $f(x)$  is a continuous function on  $\mathbb{R}$  satisfying  $f(x+y) = f(x) + f(y)$ .

1. Prove that  $f(nx) = nf(x)$  for integers  $n$ .
2. Prove that  $f(rx) = rf(x)$  for rational numbers  $r$ .
3. Prove that  $f(x) = ax$  for some constant  $a$ .

**Exercise 2.79.** Suppose  $f(x)$  is a continuous function on  $\mathbb{R}$  satisfying  $f(x+y) = f(x)f(y)$ . Prove that either  $f(x) = 0$  or  $f(x) = a^x$  for some constant  $a > 0$ .

### Left and Right Invertibility

Let  $f(x)$  be an increasing but not necessarily continuous function on  $[a, b]$ . By Exercise 2.38, the function has only countably many discontinuities. A function  $g(x)$  on  $[f(a), f(b)]$  is a *left inverse* of  $f(x)$  if  $g(f(x)) = x$ , and is a *right inverse* if  $f(g(y)) = y$ .

**Exercise 2.80.** Prove that if  $f(x)$  has a left inverse  $g(y)$ , then  $f(x)$  is strictly increasing. Moreover, a strictly increasing function  $f(x)$  on  $[a, b]$  has a unique increasing left inverse on  $[f(a), f(b)]$ . (There may be other non-increasing left inverses.)

**Exercise 2.81.** Prove that  $f(x)$  has a right inverse if and only if  $f(x)$  is continuous. Moreover, the right inverse must be strictly increasing, and the right inverse is unique if and only if  $f(x)$  is strictly increasing.

### Directed Set

A *directed set* in a set  $I$  with a relation  $i \leq j$  defined for some (ordered) pairs of elements  $i, j$  in  $I$ , satisfying the following properties

1. Reflexivity:  $i \leq i$  for any  $i \in I$ .
2. Transitivity:  $i \leq j$  and  $j \leq k$  imply  $i \leq k$ .
3. Upper bound: For any  $i, j \in I$ , there is  $k \in I$  satisfying  $i \leq k$  and  $j \leq k$ .



We also use  $i \geq j$  to mean  $j \leq i$ .

Exercise 2.82. What makes  $\mathbb{R}$  into a directed set?

1.  $x \leq y$  means that  $x$  is less than or equal to  $y$ .
2.  $x \leq y$  means that  $x$  is bigger than or equal to  $y$ .
3.  $x \leq y$  means that  $x$  is strictly less than  $y$ .

Exercise 2.83. What makes the set  $2^X$  of all subsets of  $X$  into a directed set?

1.  $A \leq B$  means that  $A \subset B$ .
2.  $A \leq B$  means that  $A \subset B$  and  $A \neq B$ .
3.  $A \leq B$  means that  $A$  and  $B$  are disjoint.
4.  $A \leq B$  means that  $A$  contains more elements than  $B$ .

Exercise 2.84. What makes the set of sequences of real numbers into a directed set?

1.  $\{x_n\} \leq \{y_n\}$  means that  $x_n \leq y_n$  for each  $n$ .
2.  $\{x_n\} \leq \{y_n\}$  means that  $x_n \leq y_n$  for sufficiently big  $n$ .
3.  $\{x_n\} \leq \{y_n\}$  means that  $x_n \leq y_n$  for some  $n$ .

Exercise 2.85. Which are directed sets?

1.  $I = \mathbb{N}$ ,  $m \leq n$  means that  $m$  is less than or equal to  $n$ .
2.  $I = \mathbb{N}$ ,  $m \leq n$  means that  $m$  is divisible by  $n$ .
3.  $I = \mathbb{Z}$ ,  $m \leq n$  means that  $m$  is divisible by  $n$ .
4.  $I = \mathbb{R} - \{a\}$ ,  $x \leq y$  means that  $|x - a|$  is no less than  $|y - a|$ .
5.  $I =$  all subspaces of a vector space,  $V \leq W$  means that  $V$  is a subspace of  $W$ .
6.  $I =$  all partitions of an interval,  $P \leq Q$  means that  $Q$  refines  $P$ .
7.  $I =$  all partitions of an interval,  $P \leq Q$  means that  $\|Q\| \leq \|P\|$ .
8.  $I =$  all partitions of an interval,  $P \leq Q$  means that  $P$  contains fewer partition points than  $Q$ .

Exercise 2.86. Suppose  $\varphi: I \rightarrow J$  is a surjective map and  $J$  is a directed set. Define a relation  $i \leq i'$  in  $I$  when  $\varphi(i) \leq \varphi(i')$ . Prove that the relation makes  $I$  into a direct set. Explain that surjection condition is necessary.

Exercise 2.87. A subset  $J$  of a directed set  $I$  is *cofinal* if for any  $i \in I$ , there is  $j \in J$  satisfying  $i \leq j$ . Prove that any cofinal subset of a directed set is a directed set.

Exercise 2.88. Suppose  $I$  and  $J$  are directed sets. What makes  $I \times J$  into a directed set?

1.  $(i, j) \leq (i', j')$  if  $i \leq i'$  and  $j \leq j'$ .
2.  $(i, j) \leq (i', j')$  if either  $i \leq i'$ , or  $i = i'$  and  $j \leq j'$  (*lexicographical order*).

### Limit over a Directed Set

A function  $f$  on a directed set  $I$  converges to a limit  $l$ , and denoted  $\lim_{I, \leq} f(i) = l$ , if for any  $\epsilon > 0$ , there is  $i_0 \in I$ , such that

$$i \geq i_0 \implies |f(i) - l| < \epsilon.$$

Exercise 2.89. Explain that the following limits are limits over some directed sets.

1. Limit of a sequence of real numbers.
2. Limit of a function at  $a$  (the function is defined near  $a$  but not necessarily at  $a$ ).
3. Right limit of a function at  $a$ .
4. Limit of a function at  $+\infty$ .
5. Riemann integral of a bounded function on a bounded interval.

Exercise 2.90. Explain that Darboux integral is the limit of the Riemann sum  $S(P, f)$  on the directed set of all partitions, with  $P \leq Q$  meaning that  $Q$  is a refinement of  $P$ .

Exercise 2.91. Suppose  $I$  has a *terminal element*  $t$ , which means that  $i \leq t$  for all  $i \in I$ . Prove that  $\lim_{I, \leq} f(i)$  always converges to  $f(t)$ .

Exercise 2.92. Show that the definition of limit is the same if  $< \epsilon$  is replaced by either  $\leq \epsilon$  or  $\leq \epsilon^2$ .

Exercise 2.93. Show that the definition of limit may not be the same if  $i \geq i_0$  is replaced by  $i > i_0$  (i.e.,  $i \geq i_0$  and  $i \neq i_0$ ). Find a suitable condition so that the definition remains the same.

Exercise 2.94. Prove the uniqueness of the value of the limit.

Exercise 2.95. Formulate and prove the arithmetic properties of the limit.

Exercise 2.96. Do we still have the order rule and the sandwich rule for limit over a directed set?

Exercise 2.97. Suppose  $\leq$  and  $\leq'$  are two relations on  $I$ , making  $I$  into two directed sets. If  $i \leq j$  implies  $i' \leq j'$ , how can you compare  $\lim_{I, \leq}$  and  $\lim_{I, \leq'}$ . Moreover, use your conclusion to compare the Riemann integral and the Darboux integral.

Exercise 2.98. Suppose  $J$  is a cofinal subset of a directed set  $I$ , defined in Example 2.87. Prove that  $\lim_{I, \leq} f(i) = l$  implies  $\lim_{J, \leq} f(i) = l$ . This generalises the limit of subsequence.

Exercise 2.99. Define the monotone property of a function on a directed set. Is it true that an increasing and bounded function always converges?

Exercise 2.100. Is it possible to define the concept of upper limit for a function on a directed set?

**Cauchy Criterion on a Directed Set**

A function  $f$  on a directed set  $I$  satisfies the *Cauchy criterion*, if for any  $\epsilon > 0$ , there is  $i_0$ , such that

$$i, j \geq i_0 \implies |f(i) - f(j)| \leq \epsilon.$$

**Exercise 2.101.** Prove that if  $\lim_{I, \leq} f(i)$  converges, then  $f$  satisfies the Cauchy criterion.

**Exercise 2.102.** Suppose a function  $f$  satisfies the Cauchy criterion. Use the following steps to prove its convergence.

1. For each natural number  $n$ , find  $i_n \in I$ , such that  $i_n \leq i_{n+1}$ , and  $i, j \geq i_n$  implies  $|f(i) - f(j)| < \frac{1}{n}$ .
2. Prove that  $\lim_{n \rightarrow \infty} f(i_n)$  converges. Let the limit be  $l$ .
3. Prove that  $\lim_{I, \leq} f(i) = l$ .



## **Chapter 3**

# **Differentiation**

### 3.1 Linear Approximation

Suppose  $P(A)$  is a problem about an object  $A$ . To solve the problem, we may consider a simpler object  $B$  that closely approximates  $A$ . Because  $B$  is simpler,  $P(B)$  is easier to solve. Moreover, because  $B$  is close to  $A$ , the (easily obtained) answer to  $P(B)$  is also pretty much the answer to  $P(A)$ .

Many real world problems can be mathematically interpreted as quantities related by complicated functions. To solve the problem, we may try to approximate the complicated functions by simple ones and solve the same problem for the simple functions.

What are the simple functions? Although the answer could be rather subjective, most people would agree that the following functions are listed from simple to complicated.

1. constant:  $1, \sqrt{2}, -\pi$ .
2. linear:  $3 + x, 4 - 5x$ .
3. quadratic:  $1 + x^2, 4 - 5x + 2x^2$ .
4. cubic:  $2x - 5x^3$ .
5. rational:  $\frac{1}{1+x}, \frac{2+x^2}{x(3-2x)}$ .
6. algebraic:  $\sqrt{x}, (1+x^{\frac{1}{2}})^{-\frac{2}{3}}$ .
7. transcendental:  $\sin x, e^x + 2\cos x$ .

What do we mean by approximating a function by a simple (class of) functions? Consider measuring certain length (say the height of a person, for example) by a ruler with only centimeters. We expect to get an approximate reading of the height with the accuracy within millimeters, or significantly smaller than the base unit of 1cm for the ruler:

$$|\text{actual length} - \text{reading from ruler}| \leq \epsilon(1\text{cm}).$$

Similarly, approximating a function  $f(x)$  at  $x_0$  by a function  $p(x)$  of some class should mean that, as  $x$  approaches  $x_0$ , the difference  $|f(x) - p(x)|$  is significantly smaller than the “base unit”  $u(x)$  for the class.

**Definition 3.1.1.** A function  $f(x)$  is approximated at  $x_0$  by a function  $p(x)$  with respect to the base unit function  $u(x) \geq 0$ , denoted  $f \sim_u p$ , if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|x - x_0| < \delta \implies |f(x) - p(x)| \leq \epsilon u(x).$$

We may also define one sided version of approximation. This is left as Exercise 3.6.

We may express the definition as

$$f(x) = p(x) + o(u(x)),$$

where  $o(u(x))$  denotes any function  $r(x)$  (in our case the *error* or *remainder*  $f(x) - p(x)$ ) satisfying the property that, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $u(x)$  may take negative value in general

$$|x - x_0| < \delta \implies |r(x)| \leq \epsilon |u(x)|.$$

In case  $u(x) \neq 0$  for  $x \neq x_0$ , this means  $r(x_0) = 0$  and  $\lim_{x \rightarrow x_0} \frac{r(x)}{u(x)} = 0$ .

For the class of constant functions  $p(x) = a$ , the base unit is  $u(x) = 1$ . The approximation of  $f(x)$  by  $p(x) = a$  at  $x_0$  means that, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|x - x_0| < \delta \implies |f(x) - a| \leq \epsilon.$$

Taking  $x = x_0$ , we get  $|f(x_0) - a| \leq \epsilon$  for any  $\epsilon > 0$ . This means exactly  $a = f(x_0)$ . On the other hand, for  $x \neq x_0$ , we get

$$0 < |x - x_0| < \delta \implies |f(x) - a| \leq \epsilon.$$

This means exactly  $\lim_{x \rightarrow x_0} f(x) = a$ . Combined with  $a = f(x_0)$ , we conclude that a function  $f(x)$  is approximated by a constant at  $x_0$  if and only if it is continuous at  $x_0$ . Moreover, the approximating constant is  $f(x_0)$ .

**Exercise 3.1.** Suppose  $f(x)$  approximated at  $x_0$  by a positive number. Prove that there is  $\delta > 0$ , such that  $f > 0$  on  $(x_0 - \delta, x_0 + \delta)$ .

**Exercise 3.2.** Define the approximation by constant at the right of  $x_0$  and explain that the concept is equivalent to the right continuity.

**Exercise 3.3.** Prove that the approximation with respect to the same unit  $u$  is an equivalence relation.

1.  $f \sim_u f$ .
2. If  $f \sim_u g$ , then  $g \sim_u f$ .
3. If  $f \sim_u g$  and  $g \sim_u h$ , then  $f \sim_u h$ .

**Exercise 3.4.** Suppose  $u(x) \leq Cv(x)$  for a constant  $C > 0$ . Prove that  $f \sim_u p$  implies  $f \sim_v p$ . This means more refined approximation (measured by  $u$ ) implies less refined approximation (measured by  $v$ ).

**Exercise 3.5.** Prove the following properties of  $o(u(x))$  ( $u(x)$  is not assumed to be non-negative).

1.  $o(u(x)) + o(u(x)) = o(u(x))$ .
2.  $o(u(x))v(x) = o(u(x)v(x))$ .
3. If  $|u(x)| \leq C|v(x)|$ , then  $o(v(x)) = o(u(x))$ .

**Exercise 3.6.** Define one sided approximation such as  $f \sim_u p$  at  $x_0^+$ . Explain the relation between the usual approximation and one sided approximations. Moreover, what is the meaning of one sided constant approximation?

## Differentiability

The approximation by constant functions is too crude for solving most problems. As the next simplest, we need to consider the approximation by *linear functions*  $p(x) = A + Bx$ . For the convenience of discussion, we introduce

$$\Delta x = x - x_0,$$

( $\Delta$  is the Greek alphabet for  $D$ , used here for the *Difference*) and rewrite

$$p(x) = A + Bx = A + B(x_0 + \Delta x) = (A + Bx_0) + B\Delta x = a + b\Delta x.$$

What is the base unit of  $a + b\Delta x$  as  $x$  approaches  $x_0$ ? The base unit of the constant term  $a$  is 1. The base unit of the difference term  $b\Delta x$  is  $|\Delta x|$ , which is very small compared with the unit 1. Therefore the base unit for the linear function  $a + b\Delta x$  is  $u(x) = |\Delta x|$ . The discussion may be compared with the expression  $am + bcm$  ( $a$  meters and  $b$  centimeters). Since 1cm is much smaller than 1m, the base unit for  $am + bcm$  is 1cm.

**Definition 3.1.2.** A function  $f(x)$  is *differentiable* at  $x_0$  if it is approximated by a linear function  $a + b\Delta x$ . In other words, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|\Delta x| = |x - x_0| < \delta \implies |f(x) - a - b\Delta x| \leq \epsilon |\Delta x|. \quad (3.1.1)$$

We may express the linear approximation as

$$f(x) = a + b\Delta x + o(\Delta x),$$

where  $o(\Delta x)$  means any function  $r(x)$  satisfying  $\lim_{x \rightarrow x_0} \frac{r(x)}{\Delta x} = 0$ . Taking  $x = x_0$  in (3.1.1), we get  $a = f(x_0)$ . Then

$$\Delta f = f(x) - a = f(x) - f(x_0)$$

is the change of the function caused by the change  $\Delta x$  of the variable, and the differentiability means

$$\Delta f = b\Delta x + o(\Delta x).$$

In other words, the scaling  $b\Delta x$  of the change of variable is the linear approximation of the change of function. The viewpoint leads to the *differential* of the function at  $x_0$

$$df = b dx.$$

Note that the symbols  $df$  and  $dx$  have not yet been specified as numerical quantities. Thus  $b dx$  should be, at least for the moment, considered as an *integrated*



notation instead of the product of two quantities. On the other hand, the notation is motivated from  $b\Delta x$ , which was indeed a product of two numbers. So it is allowed to add two differentials and to multiply numbers to differentials. In more advanced mathematics (see Section 14.3), the differential symbols will indeed be defined as quantities in some linear approximation space. However, one has to be careful in multiplying differentials together because this is not a valid operation within linear spaces. Moreover, dividing differentials is also not valid.

Linear approximation is more refined than constant approximation. We expect more refined approximation to imply less refined approximation.

**Proposition 3.1.3.** *If a function is differentiable at  $x_0$ , then it is continuous at  $x_0$ .*

*Proof.* Taking  $\epsilon = 1$  in the definition of differentiability, we have  $\delta > 0$ , such that

$$\begin{aligned} |\Delta x| = |x - x_0| < \delta &\implies |f(x) - f(x_0) - b\Delta x| \leq |\Delta x| \\ &\implies |f(x) - f(x_0)| \leq (|b| + 1)|\Delta x|. \end{aligned}$$

Then for any  $\epsilon > 0$ , we get

$$|x - x_0| < \max \left\{ \delta, \frac{\epsilon}{|b| + 1} \right\} \implies |f(x) - f(x_0)| \leq (|b| + 1)|x - x_0| \leq \epsilon.$$

This proves the continuity at  $x_0$ . □

**Example 3.1.1.** Since the function  $2x + 3$  is already linear, its linear approximation is itself. Expressed in the form  $a + b\Delta x$ , the linear approximation at  $x_0 = 0$  is  $3 + 2x$ , the linear approximation at  $x_0 = 1$  is  $5 + 2(x - 1)$ , and the linear approximation at  $x_0 = -1$  is  $1 + 2(x + 1)$ .

**Example 3.1.2.** To find the linear approximation of  $x^2$  at  $x_0 = 1$ , we rewrite the function in terms of  $\Delta x = x - 1$  near 1.

$$x^2 = (1 + \Delta x)^2 = 1 + 2\Delta x + \Delta x^2.$$

Then

$$|\Delta x| < \delta = \epsilon \implies |x^2 - 1 - 2\Delta x| = |\Delta x|^2 \leq \epsilon|\Delta x|.$$

This shows that  $1 + 2\Delta x$  is the linear approximation of  $x^2$  at 1.

**Exercise 3.7.** Find the differentiability of a general linear function  $Ax + B$ .

**Exercise 3.8.** Show that  $x^2$  is linearly approximated by  $x_0^2 + 2x_0\Delta x$  at  $x_0$ . What about the linear approximation of  $x^3$ ?

**Example 3.1.3.** We study the differentiability of  $|x|^p$ ,  $p > 0$ , at  $x_0 = 0$ .

If  $p > 1$ , then we can arrange to have  $|x|^p \leq \epsilon|x|$  for sufficiently small  $x$ . Specifically, we have

$$|x - 0| = |x| < \epsilon^{\frac{1}{p-1}} \implies ||x|^p - 0 - 0x| = |x|^{p-1}|x| \leq \epsilon|x|.$$

This shows that  $f(x)$  is differentiable at 0, with linear approximation  $0 + 0x = 0$ . In fact, the same argument works as long as  $|f(x)| \leq C|x|^p$ ,  $p > 1$ . For example, the function

$$x^2 D(x) = \begin{cases} x^2, & \text{if } x \text{ is rational,} \\ 0, & \text{if } x \text{ is irrational,} \end{cases}$$

is differentiable at 0, with trivial linear approximation.

If  $0 < p \leq 1$ , then differentiability at 0 means that, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|x| < \delta \implies ||x|^p - a - bx| \leq \epsilon|x|.$$

By taking  $x = 0$ , the implication says  $|a| \leq 0$ . Therefore  $a = 0$  and we get (we already discussed  $|x| = 0$ )

$$0 < |x| < \delta \implies ||x|^p - bx| \leq \epsilon|x| \iff \left| \frac{|x|^p}{x} - b \right| \leq \epsilon.$$

This means exactly that  $b = \lim_{x \rightarrow 0} \frac{|x|^p}{x}$  converges. Since the limit diverges for  $0 < p \leq 1$ , we conclude that  $|x|^p$  is not differentiable at 0.

**Example 3.1.4.** Let  $p, q > 0$  and

$$f(x) = \begin{cases} x^p, & \text{if } x \geq 0, \\ -(-x)^q, & \text{if } x < 0. \end{cases}$$

We have  $f(x) = |x|^p$  on the right of 0 and  $f(x) = |x|^q$  on the left of 0. We may split the definition of linear approximation into the right and left. We get the overall differentiability only if two sides have the same linear approximation.

For  $p > 1$ , we may restrict the discussion of Example 3.1.3 to the right of 0 and find that  $f(x) = |x|^p$  is *right differentiable* at 0 with 0 as the right linear approximation. For  $p = 1$ ,  $f(x) = |x|^p = x$  is a linear function and is therefore right differentiable, with  $x$  as the right linear approximation. For  $0 < p < 1$ , the same argument as in Example 3.1.3 shows that  $f(x)$  is not right differentiable.

We have  $f(x) = -|x|^q$  on the left of 0. We can similarly discuss the left differentiability. By comparing the two sides, we find that  $f(x)$  is differentiable at 0 if and only if both  $p, q > 1$  or  $p = q = 1$ .

**Exercise 3.9.** Define one sided differentiability. Then prove that a function is differentiable at  $x_0$  if and only if it is left and right differentiable at  $x_0$ , and the left and right linear approximations are the same.

**Exercise 3.10.** Prove that right differentiability implies right continuity.

**Exercise 3.11.** Study the differentiability of function at  $x_0 = 0$

$$f(x) = \begin{cases} ax^p, & \text{if } x \geq 0, \\ b(-x)^q, & \text{if } x < 0. \end{cases}$$

**Exercise 3.12.** Split the differentiability into rational and irrational parts. Then study the differentiability of function at  $x_0 = 0$

$$f(x) = \begin{cases} |x|^p, & \text{if } x \text{ is rational,} \\ |x|^q, & \text{if } x \text{ is irrational.} \end{cases}$$

**Exercise 3.13.** Prove the sandwich rule for linear approximation: If  $f(x) \leq g(x) \leq h(x)$ , and both  $f(x)$  and  $h(x)$  are approximated by the same linear function  $a + b\Delta x$  at  $x_0$ , then  $g(x)$  is approximated by the linear function  $a + b\Delta x$ .

**Exercise 3.14.** Can you state and prove a general approximation version of the sandwich rule?

## Derivative

We have computed the constant term  $a = f(x_0)$  of the linear approximation. For  $x \neq x_0$ , the condition (3.1.1) becomes

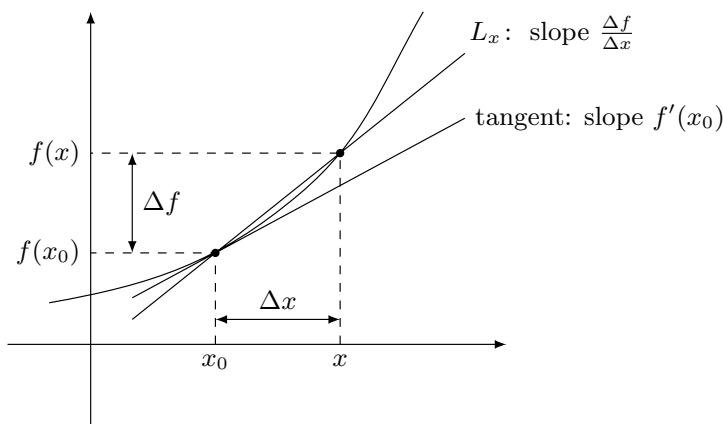
$$0 < |\Delta x| = |x - x_0| < \delta \implies \left| \frac{f(x) - f(x_0)}{x - x_0} - b \right| \leq \epsilon. \quad (3.1.2)$$

This shows how to compute the coefficient  $b$  of the first order term.

**Definition 3.1.4.** The *derivative* of a function  $f(x)$  at  $x_0$  is

$$f'(x_0) = \frac{df(x_0)}{dx} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x}.$$

Corresponding to the left and right approximations (see Exercises 3.6 and 3.9), we also have one sided derivatives. See Example 3.1.8 and Exercises 3.20, 3.22).



**Figure 3.1.1.** Linear approximation and derivative of  $f(x)$  at  $x_0$ .

We already saw that the differentiability implies the existence of derivative. Conversely, if the derivative exists, then we have the implication (3.1.2), which is

the same as (3.1.1) with  $a = f(x_0)$  and  $0 < |x - x_0| < \delta$ . Since (3.1.1) always holds with  $a = f(x_0)$  and  $x = x_0$ , we conclude that the condition for differentiability holds with  $a = f(x_0)$  and  $|x - x_0| < \delta$ .

**Proposition 3.1.5.** *A function  $f(x)$  is differentiable at  $x_0$  if and only if it has derivative at  $x_0$ . Moreover, the linear approximation is  $f(x_0) + f'(x_0)\Delta x$ .*

Geometrically, the quotient  $\frac{\Delta f}{\Delta x}$  is the slope of the straight line connecting  $(x_0, f(x_0))$  to a nearby point  $(x, f(x))$ . As the limit of this slope, the derivative  $f'(x_0)$  is the slope of the tangent line at  $x_0$ . The formula for the tangent line is  $y = f(x_0) + f'(x_0)\Delta x$ , which is the linear approximation function.

We emphasize that, although the existence of linear approximation is equivalent to the existence of derivative, the two play different roles. Linear approximation is the motivation and the concept. Derivative, as the coefficient of the first order term, is merely the computation of the concept. Therefore linear approximation is much more important in understanding the essence of calculus. As a matter of fact, for multivariable functions, linear approximations may be similarly computed by *partial derivatives*. However, the existence of partial derivatives does not necessarily imply the existence of linear approximation.

Mathematical concepts are always derived from common sense. The formula for computing a concept is obtained only after analyzing the common sense. Never equate the formula with the concept itself!

**Example 3.1.5.** Example 3.1.1 shows that the linear approximation of a linear function  $A + Bx$  is itself. The coefficient of the first order term is then the derivative  $(A + Bx)' = B$ .

**Example 3.1.6.** For a quadratic function  $f(x) = A + Bx + Cx^2$ , we have

$$\Delta f = f(x_0 + \Delta x) - f(x_0) = (B + 2Cx_0)\Delta x + C\Delta x^2 = (B + 2Cx_0)\Delta x + o(\Delta x).$$

Therefore the function is differentiable at  $x_0$ , with  $f'(x_0) = B + 2Cx_0$ . We usually write  $f'(x) = B + 2Cx$ . In terms of differential, we write  $df = (B + 2Cx)dx$ .

**Example 3.1.7.** For  $f(x) = \sin x$ , the limit (2.2.10) gives

$$f'(0) = \lim_{x \rightarrow 0} \frac{f(x) - f(0)}{x} = \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1.$$

Therefore the sine function is differentiable at 0, with  $f(0) + f'(0)(x - 0) = x$  as the linear approximation.

**Example 3.1.8.** The function  $f(x) = |x|$  has no derivative because

$$\lim_{x \rightarrow 0} \frac{f(x) - f(0)}{x} = \lim_{x \rightarrow 0} \frac{|x|}{x}$$

diverges. Therefore  $|x|$  is not differentiable at 0. The conclusion is consistent with Example 3.1.3.

We notice that the *right derivative*

$$f'_+(0) = \lim_{x \rightarrow 0^+} \frac{f(x) - f(0)}{x} = \lim_{x \rightarrow 0^+} \frac{x}{x} = 1$$

and the *left derivative*

$$f'_-(0) = \lim_{x \rightarrow 0^-} \frac{f(x) - f(0)}{x} = \lim_{x \rightarrow 0^-} \frac{-x}{x} = -1$$

exist. The derivative does not exist because the two one sided derivatives have different values.

**Exercise 3.15.** Prove the uniqueness of the linear approximation: If both  $a + b\Delta x$  and  $a' + b'\Delta x$  are linear approximations of  $f(x)$  at  $x_0$ , then  $a = a'$  and  $b = b'$ .

**Exercise 3.16.** Find the derivative and the differential for the cubic function  $f(x) = x^3$  by computing  $\Delta f = f(x_0 + \Delta x) - f(x_0)$ .

**Exercise 3.17.** Prove that  $f(x)$  is differentiable at  $x_0$  if and only if  $f(x) = f(x_0) + (x - x_0)g(x)$  for a function  $g(x)$  continuous at  $x_0$ . Moreover, the linear approximation is  $f(x_0) + g(x_0)(x - x_0)$ .

**Exercise 3.18.** Rephrase the sandwich rule in Exercise 3.13 in terms of the derivative.

**Exercise 3.19.** Suppose  $f(x_0) = g(x_0) = 0$ ,  $f(x)$  and  $g(x)$  are continuous at  $x_0$ , and  $\lim_{x \rightarrow x_0} \frac{g(x)}{f(x)} = 1$ . Prove that  $f(x)$  is differentiable at  $x_0$  if and only if  $g(x)$  is differentiable at  $x_0$ . Moreover, the linear approximations of the two functions are the same.

**Exercise 3.20.** Define the right derivative  $f'_+(x_0)$  and the left derivative  $f'_-(x_0)$ . Then show that the existence of one sided derivative is equivalent to one sided differentiability in Exercise 3.9.

**Exercise 3.21.** What is the relation between one sided derivative and one sided continuity?

**Exercise 3.22.** What is the relation between the usual two sided derivative and the one sided derivatives?

**Exercise 3.23.** Determine the differentiability of  $|x^3(x - 1)(x - 2)^2|$ .

**Exercise 3.24.** For  $p > 0$ , compute the right derivative of the power function  $x^p$  at 0.

**Exercise 3.25.** Study the right differentiability of the function  $\begin{cases} x^x, & \text{if } x > 0 \\ 1, & \text{if } x = 0 \end{cases}$  at 0.

**Exercise 3.26.** Determine the differentiability of Thomae's function in Example 2.3.2.

**Exercise 3.27.** Suppose  $f(0) = 0$  and  $f(x) \geq |x|$ . Show that  $f(x)$  is not differentiable at 0. What if  $f(x) \leq |x|$ ?

Exercise 3.28. Suppose  $f(x)$  is differentiable on a bounded interval  $(a-\epsilon, b+\epsilon)$ . If  $f(x) = 0$  for infinitely many  $x \in [a, b]$ , prove that there is  $c \in [a, b]$ , such that  $f(c) = 0$  and  $f'(c) = 0$ .

Exercise 3.29. Suppose  $f(x)$  is differentiable at  $x_0$  and  $f(x_0) = 1$ . Find  $\lim_{t \rightarrow 0} f(x_0 + t)^{\frac{1}{t}}$ .

Exercise 3.30. For a continuous function  $f(x)$ , define

$$g(x) = \begin{cases} f(x), & \text{if } x \text{ is rational,} \\ -f(x), & \text{if } x \text{ is irrational.} \end{cases}$$

Find the necessary and sufficient condition for  $g$  to be continuous at  $x_0$ , and the condition for  $g$  to be differentiable at  $x_0$ .

## Basic Derivatives

For  $x_0 \neq 0$  and  $t = \frac{\Delta x}{x_0}$ , the limit (2.5.5) implies

$$\left. \frac{dx^p}{dx} \right|_{x=x_0} = \lim_{\Delta x \rightarrow 0} \frac{(x_0 + \Delta x)^p - x_0^p}{\Delta x} = \lim_{t \rightarrow 0} \frac{x_0^p}{x_0} \cdot \frac{(1+t)^p - 1}{t} = px_0^{p-1}.$$

We denote the result as  $(x^p)' = px^{p-1}$ ,  $dx^p = px^{p-1}dx$ .

The limit (2.5.4) implies

$$\left. \frac{da^x}{dx} \right|_{x=x_0} = \lim_{\Delta x \rightarrow 0} \frac{a^{x_0 + \Delta x} - a^{x_0}}{\Delta x} = \lim_{\Delta x \rightarrow 0} a^{x_0} \frac{a^{\Delta x} - 1}{\Delta x} = a^{x_0} \log a.$$

We denote the result as  $(a^x)' = a^x \log a$ ,  $da^x = a^x(\log a)dx$ .

For  $x_0 > 0$  and  $t = \frac{\Delta x}{x_0}$ , the limit (2.5.2) implies

$$\begin{aligned} \left. \frac{d \log x}{dx} \right|_{x=x_0} &= \lim_{\Delta x \rightarrow 0} \frac{\log(x_0 + \Delta x) - \log x_0}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\log(1 + \frac{\Delta x}{x_0})}{\Delta x} \\ &= \lim_{t \rightarrow 0} \frac{1}{x_0} \cdot \frac{\log(1+t)}{t} = \frac{1}{x_0}. \end{aligned}$$

We denote the result as  $(\log x)' = \frac{1}{x}$ ,  $d \log x = \frac{1}{x}dx$ .

The limits (2.2.10) and (2.2.12) imply

$$\begin{aligned} \left. \frac{d \sin x}{dx} \right|_{x=x_0} &= \lim_{\Delta x \rightarrow 0} \frac{\sin(x_0 + \Delta x) - \sin x_0}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{\sin x_0 \cos \Delta x + \cos x_0 \sin \Delta x - \sin x_0}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \left( -\frac{1 - \cos \Delta x}{\Delta x} \sin x_0 + \frac{\sin \Delta x}{\Delta x} \cos x_0 \right) \\ &= -0 \cdot \sin x_0 + 1 \cdot \cos x_0 = \cos x_0. \end{aligned}$$

We denote the result as  $(\sin x)' = \cos x$ ,  $d \sin x = \cos x dx$ . Similarly, we have  $(\cos x)' = -\sin x$ ,  $d \cos x = -\sin x dx$  and  $(\tan x)' = \frac{1}{\cos^2 x} = \sec^2 x$ ,  $d \tan x = \sec^2 x dx$ .

Exercise 3.31. Find the derivatives of  $\cos x$  and  $\tan x$  by definition.

## 3.2 Computation

### Combination of Linear Approximation

Suppose we have linear approximations at  $x_0$

$$\begin{aligned} f(x) &\sim_{|\Delta x|} p(x) = f(x_0) + f'(x_0)\Delta x, \\ g(x) &\sim_{|\Delta x|} q(x) = g(x_0) + g'(x_0)\Delta x. \end{aligned}$$

Then we expect linear approximation

$$f(x) + g(x) \sim_{|\Delta x|} p(x) + q(x) = (f(x_0) + g(x_0)) + (f'(x_0) + g'(x_0))\Delta x.$$

In particular, this implies that the derivative  $(f + g)'(x_0)$  is the coefficient  $f'(x_0) + g'(x_0)$  of  $\Delta x$ . We express the fact as

$$(f + g)' = f' + g', \quad d(f + g) = df + dg.$$

Similarly, we expect

$$\begin{aligned} f(x)g(x) &\sim_{|\Delta x|} p(x)q(x) \\ &= f(x_0)g(x_0) + (f'(x_0)g(x_0) + f(x_0)g'(x_0))\Delta x + f'(x_0)g'(x_0)\Delta x^2. \end{aligned}$$

Although  $p(x)q(x)$  is not linear, it should be further approximated by the linear function

$$r(x) = f(x_0)g(x_0) + (f'(x_0)g(x_0) + f(x_0)g'(x_0))\Delta x.$$

Then we should have  $f(x)g(x) \sim_{|\Delta x|} r(x)$ . In particular,  $(fg)'(x_0)$  is the coefficient  $f'(x_0)g(x_0) + f(x_0)g'(x_0)$  of  $\Delta x$ . We express the fact as the *Leibniz*<sup>11</sup> rule

$$(fg)' = f'g + fg', \quad d(fg) = gdf + fdg.$$

For the composition  $g(f(x))$ , the functions  $f(x)$  and  $g(y)$  are approximated by linear functions at  $x_0$  and  $y_0 = f(x_0)$

$$\begin{aligned} f(x) &\sim_{|\Delta x|} p(x) = f(x_0) + f'(x_0)\Delta x, \\ g(y) &\sim_{|\Delta y|} q(y) = g(y_0) + g'(y_0)\Delta y. \end{aligned}$$

---

<sup>11</sup>Gottfried Wilhelm von Leibniz, born 1646 in Leipzig, Saxony (Germany), died 1716 in Hanover (Germany). Leibniz was a great scholar who contributed to almost all the subjects in the human knowledge of his time. He invented calculus independent of Newton. He also invented the binary system, the foundation of modern computer system. He was, along with René Descartes and Baruch Spinoza, one of the three greatest 17th-century rationalists. Leibniz was perhaps the first major European intellect who got seriously interested in Chinese civilization. His fascination of I Ching may be related to his invention of binary system.

Then we expect the composition to be approximated by the composition of linear functions

$$g(f(x)) \sim_{|\Delta x|} q(p(x)) = g(y_0) + g'(y_0)f'(x_0)\Delta x.$$

This implies that the derivative  $(g \circ f)'(x_0)$  of the composition is the coefficient  $g'(y_0)f'(x_0) = g'(f(x_0))f'(x_0)$  of  $\Delta x$ . We express the fact as the *chain rule*

$$(g \circ f)' = (g' \circ f)f', \quad d(g \circ f) = (g' \circ f)df.$$

**Proposition 3.2.1.** *The sum, the product, the composition of differentiable functions are differentiable, with the respective linear approximations obtained by the sum, the linear truncation of product, and the composition of linear approximations.*

The sum property

$$f \sim p, g \sim q \implies f + g \sim p + q$$

is parallel to the property  $\lim(x_n + y_n) = \lim x_n + \lim y_n$  in Proposition 1.2.3

$$x_n \sim l, y_n \sim k \implies x_n + y_n \sim l + k.$$

Therefore the formula  $(f + g)' = f' + g'$  can be proved by copying the earlier proof. The same remark applies to the product and the composition.

*Proof.* First consider the sum. For any  $\epsilon_1 > 0$ , there are  $\delta_1, \delta_2 > 0$ , such that

$$|\Delta x| < \delta_1 \implies |f(x) - p(x)| \leq \epsilon_1 |\Delta x|, \quad (3.2.1)$$

$$|\Delta x| < \delta_2 \implies |g(x) - q(x)| \leq \epsilon_1 |\Delta x|. \quad (3.2.2)$$

This implies

$$|\Delta x| < \min\{\delta_1, \delta_2\} \implies |(f(x) + g(x)) - (p(x) + q(x))| \leq 2\epsilon_1 |\Delta x|. \quad (3.2.3)$$

Therefore for any  $\epsilon > 0$ , we may take  $\epsilon_1 = \frac{\epsilon}{2}$  and find  $\delta_1, \delta_2 > 0$ , such that (3.2.1) and (3.2.2) hold. Then (3.2.3) holds and becomes

$$|\Delta x| < \min\{\delta_1, \delta_2\} \implies |(f(x) + g(x)) - (p(x) + q(x))| \leq \epsilon |\Delta x|.$$

This completes the proof that  $p(x) + q(x)$  is the linear approximation of  $f(x) + g(x)$ .

Now consider the product. For  $|\Delta x| < \min\{\delta_1, \delta_2\}$ , (3.2.1) and (3.2.2) imply

$$\begin{aligned} |f(x)g(x) - p(x)q(x)| &\leq |f(x)g(x) - p(x)g(x)| + |p(x)g(x) - p(x)q(x)| \\ &\leq \epsilon_1 |\Delta x| |g(x)| + \epsilon_1 |p(x)| |\Delta x| \\ &\leq \epsilon_1 (|g(x)| + |p(x)|) |\Delta x|. \end{aligned}$$

This further implies

$$\begin{aligned} |f(x)g(x) - r(x)| &\leq |f(x)g(x) - p(x)q(x)| + |f'(x_0)g'(x_0)\Delta x^2| \\ &\leq [\epsilon_1 (|g(x)| + |p(x)|) + |f'(x_0)g'(x_0)\Delta x|] |\Delta x|. \end{aligned}$$



Since  $|g(x)|$  and  $|p(x)|$  are continuous at  $x_0$ , they are bounded near  $x_0$  by Proposition 2.1.6. Therefore for any  $\epsilon > 0$ , it is not difficult to find  $\delta_3 > 0$  and  $\epsilon_1 > 0$ , such that

$$\epsilon_1(|g(x)| + |p(x)|) + |f'(x_0)g'(x_0)|\delta_3 \leq \epsilon.$$

Next for this  $\epsilon_1$ , we may find  $\delta_1, \delta_2 > 0$ , such that (3.2.1) and (3.2.2) hold. Then  $|\Delta x| < \min\{\delta_1, \delta_2, \delta_3\}$  implies

$$|f(x)g(x) - r(x)| \leq [\epsilon_1(|g(x)| + |p(x)|) + |f'(x_0)g'(x_0)|\delta_3]|\Delta x| \leq \epsilon|\Delta x|.$$

This completes the proof that the linear truncation  $r(x)$  of  $p(x)q(x)$  is the linear approximation of  $f(x)g(x)$ .

Finally consider the composition. For any  $\epsilon_1, \epsilon_2 > 0$ , there are  $\delta_1, \delta_2 > 0$ , such that

$$|\Delta x| = |x - x_0| < \delta_1 \implies |f(x) - p(x)| \leq \epsilon_1|\Delta x|, \quad (3.2.4)$$

$$|\Delta y| = |y - y_0| < \delta_2 \implies |g(y) - q(y)| \leq \epsilon_2|\Delta y|. \quad (3.2.5)$$

Then for  $y = f(x)$ , we have

$$\begin{aligned} |\Delta x| < \delta_1 \implies |\Delta y| &= |f(x) - f(x_0)| \leq |f(x) - p(x)| + |f'(x_0)\Delta x| \\ &\leq (\epsilon_1 + |f'(x_0)|)|\Delta x| < (\epsilon_1 + |f'(x_0)|)\delta_1, \end{aligned} \quad (3.2.6)$$

and

$$\begin{aligned} |\Delta x| < \delta_1, |\Delta y| < \delta_2 \implies &|g(f(x)) - q(p(x))| \\ &\leq |g(f(x)) - q(f(x))| + |q(f(x)) - q(p(x))| \\ &= |g(y) - q(y)| + |g'(x_0)||f(x) - p(x)| \\ &\leq \epsilon_2|\Delta y| + |g'(x_0)|\epsilon_1|\Delta x| \\ &\leq [\epsilon_2(\epsilon_1 + |f'(x_0)|) + \epsilon_1|g'(x_0)|]|\Delta x|. \end{aligned} \quad (3.2.7)$$

Suppose for any  $\epsilon > 0$ , we can find suitable  $\delta_1, \delta_2, \epsilon_1, \epsilon_2$ , such that (3.2.12), (3.2.13) hold, and

$$(\epsilon_1 + |f'(x_0)|)\delta_1 \leq \delta_2, \quad (3.2.8)$$

$$\epsilon_2(\epsilon_1 + |f'(x_0)|) + \epsilon_1|g'(x_0)| \leq \epsilon. \quad (3.2.9)$$

Then (3.2.6) tells us that  $|\Delta x| < \delta_1$  implies  $|\Delta y| < \delta_2$ , and (3.2.7) becomes

$$\begin{aligned} |\Delta x| < \delta_1 \implies &|g(f(x)) - q(p(x))| \\ &\leq [\epsilon_2(\epsilon_1 + |f'(x_0)|) + \epsilon_1|g'(x_0)|]|\Delta x| \leq \epsilon|\Delta x|. \end{aligned}$$

This would prove that  $q(p(x))$  is the linear approximation of  $g(f(x))$ .

It remains to find  $\delta_1, \delta_2, \epsilon_1, \epsilon_2$  such that (3.2.12), (3.2.13), (3.2.16) and (3.2.9) hold. For any  $\epsilon > 0$ , we first find  $\epsilon_1, \epsilon_2 > 0$  satisfying (3.2.9). For this  $\epsilon_2 > 0$ , we find  $\delta_2 > 0$ , such that the implication (3.2.13) holds. Then for the  $\epsilon_1 > 0$  we already found, it is easy to find  $\delta_1 > 0$  such that (3.2.12) and (3.2.16) hold.  $\square$

**Exercise 3.32.** A function is *odd* if  $f(-x) = -f(x)$ . It is *even* if  $f(-x) = f(x)$ . A function is *periodic* with period  $p$  if  $f(x + p) = f(x)$ . Prove that the derivative of an odd, even, or periodic function is respectively even, odd, periodic.

**Exercise 3.33.** Suppose  $f(x)$  is differentiable. Find the derivative of  $\frac{1}{f(x)}$ . Can you explain the derivative in terms of linear approximation?

**Exercise 3.34.** Find the formula for the derivative of  $\frac{f(x)}{g(x)}$ .

**Exercise 3.35.** Find the derivatives of  $\tan x$  and  $\sec x$  by using the derivatives of  $\sin x$  and  $\cos x$ .

**Exercise 3.36.** Prove  $(f + g)'_+(x) = f'_+(x) + g'_+(x)$  and  $(fg)'_+(x) = f'_+(x)g(x) + f(x)g'_+(x)$  for the right derivative.

**Exercise 3.37.** Suppose  $f(x)$  and  $g(y)$  are right differentiable at  $x_0$  and  $y_0 = f(x_0)$ . Prove that under one of the following conditions, the composition  $g(f(x))$  is right differentiable at  $x_0$ , and we have the chain rule  $(g \circ f)'_+(x_0) = g'_+(y_0)f'_+(x_0)$ .

1.  $f(x) \geq f(x_0)$  for  $x \geq x_0$ .
2.  $g(y)$  is (two sided) differentiable at  $y_0$ .

Note that by the proof of Proposition 3.3.1, the first condition is satisfied if  $f'_+(x_0) > 0$ . Can you find the other chain rules for one sided derivatives?

## Inverse Linear Approximation

Suppose we have linear approximation

$$f(x) \sim_{|\Delta x|} p(x) = a + b\Delta x = y_0 + b(x - x_0), \quad a = f(x_0) = y_0, \quad b = f'(x_0).$$

If  $f$  is invertible, then the inverse function should be approximated by the inverse linear function

$$f^{-1}(y) \sim_{|\Delta y|} p^{-1}(y) = x_0 + b^{-1}(y - y_0) = x_0 + b^{-1}\Delta y.$$

This suggests that  $(f^{-1})'(y_0) = b^{-1} = \frac{1}{f'(x_0)}$ .

**Proposition 3.2.2.** Suppose a continuous function  $f(x)$  is invertible near  $x_0$  and is differentiable at  $x_0$ . If  $f'(x_0) \neq 0$ , then the inverse function is also differentiable at  $y_0 = f(x_0)$ , with

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}.$$

The derivatives for the trigonometric functions give the derivatives of the inverse trigonometric functions. Let  $y = \arcsin x$ . Then  $x = \sin y$ , and

$$(\arcsin x)' = \frac{1}{(\sin y)'} = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - \sin^2 y}} = \frac{1}{\sqrt{1 - x^2}}.$$

Note that the third equality makes use of  $-\frac{\pi}{2} \leq y \leq \frac{\pi}{2}$ , so that  $\cos y \geq 0$ . The derivative of the inverse cosine can be derived similarly, or from (2.5.1). For the derivative of the inverse tangent, let  $y = \arctan x$ . Then  $x = \tan y$ , and

$$(\arctan x)' = \frac{1}{(\tan y)'} = \frac{1}{\sec^2 y} = \frac{1}{1 + \tan^2 y} = \frac{1}{1 + x^2}.$$

*Proof.* By the set up before the proposition, for any  $\epsilon' > 0$ , there is  $\delta' > 0$ , such that

$$|\Delta x| < \delta' \implies |f(x) - p(x)| \leq \epsilon' |\Delta x|. \quad (3.2.10)$$

By

$$\begin{aligned} f(x) - p(x) &= y - (y_0 + b(x - x_0)) = \Delta y - b\Delta x, \\ f^{-1}(y) - p^{-1}(y) &= x - (x_0 + b^{-1}(y - y_0)) = -b^{-1}(\Delta y - b\Delta x), \end{aligned}$$

the implication (3.2.10) is the same as

$$|\Delta x| < \delta' \implies |f^{-1}(y) - p^{-1}(y)| \leq |b|^{-1} \epsilon' |\Delta x|.$$

If we can achieve

$$|\Delta y| < \delta \implies |\Delta x| < \delta' \implies |b|^{-1} \epsilon' |\Delta x| \leq \epsilon |\Delta y|, \quad (3.2.11)$$

then we get

$$|\Delta y| < \delta \implies |f^{-1}(y) - p^{-1}(y)| \leq \epsilon |\Delta y|.$$

This means that  $p^{-1}(y)$  is the linear approximation of  $f^{-1}(y)$ .

Since  $b \neq 0$ , we may additionally assume  $\epsilon' < \frac{|b|}{2}$ . Then by (3.2.10),

$$\begin{aligned} |\Delta x| < \delta' &\implies |b\Delta x| - |\Delta y| \leq |\Delta y - b\Delta x| \leq \epsilon' |\Delta x| \leq \frac{|b|}{2} |\Delta x| \\ &\implies |\Delta x| \leq 2|b|^{-1} |\Delta y|. \end{aligned}$$

Therefore the second implication in (3.2.11) can be achieved if  $2|b|^{-2} \epsilon' = \epsilon$ .

Now for any  $\epsilon > 0$ , take  $\epsilon' = \frac{1}{2}|b|^2 \epsilon$ , so that the second implication in (3.2.11) holds. Then for this  $\epsilon'$ , we find  $\delta'$ , such that (3.2.10) holds. Now for this  $\delta'$ , by making use of the continuity of the inverse function (see Theorem 2.5.3), we can find  $\delta > 0$ , such that

$$|\Delta y| = |y - y_0| < \delta \implies |\Delta x| = |x - x_0| = |f^{-1}(y) - f^{-1}(y_0)| < \delta'.$$

This is the first implication in (3.2.11). This completes the proof.  $\square$

**Exercise 3.38.** Find the derivatives of  $\arccos x$ ,  $\operatorname{arcsec} x$ .

**Exercise 3.39.** Suppose  $f(x)$  is invertible near  $x_0$  and is differentiable at  $x_0$ . Prove that if the inverse function is differentiable at  $y_0 = f(x_0)$ , then  $f'(x_0) \neq 0$ . This is the “conditional” converse of Proposition 3.2.2.

**Exercise 3.40.** Proposition 3.2.2 can also be proved by computing the derivatives directly. Specifically, prove that one of the limits

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}, \quad (f^{-1})'(y_0) = \lim_{y \rightarrow y_0} \frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0},$$

converges if and only if the other one converges. Moreover, the limits are related by  $(f^{-1})'(y_0) = \frac{1}{f'(x_0)}$  when they converge.

**Exercise 3.41.** Consider the function

$$f(x) = \begin{cases} \frac{n}{n^2 + 1}, & \text{if } x = \frac{1}{n}, n \in \mathbb{N}, \\ x, & \text{otherwise.} \end{cases}$$

Verify that  $f'(0) = 1$  but  $f(x)$  is not one-to-one. This shows that the invertibility condition is necessary in Proposition 3.2.2. In particular,  $f'(x_0) \neq 0$  does not necessarily imply the invertibility of the function near  $x_0$ .

**Exercise 3.42.** State the one sided derivative version of Proposition 3.2.2.

## Combination of General Approximation

Proposition 3.2.1 is similar to Propositions 1.2.2, 1.2.3, 2.1.6, 2.1.7. The proofs are also similar. This suggests more general principles about approximations.

The following is the general statement about sum of approximations.

**Proposition 3.2.3.** *If  $f \sim_u p$  and  $g \sim_u q$  at  $x_0$ , then  $af + bg \sim_u ap + bq$  at  $x_0$ .*

*Proof.* For any  $\epsilon > 0$ , we apply the assumptions  $f \sim_u p$  and  $g \sim_u q$  to  $\frac{\epsilon}{|a|+|b|} > 0$ . Then there is  $\delta > 0$ , such that  $|\Delta x| < \delta$  implies

$$|f(x) - p(x)| \leq \frac{\epsilon}{|a| + |b|} u(x), \quad |g(x) - q(x)| \leq \frac{\epsilon}{|a| + |b|} u(x).$$

Therefore  $|\Delta x| < \delta$  further implies

$$|(af(x) + bg(x)) - (ap(x) + bq(x))| \leq |a||f(x) - p(x)| + |b||g(x) - q(x)| \leq \epsilon u(x).$$

This completes the proof of  $af + bg \sim_u ap + bq$ .  $\square$

The following is the general statement about product of approximations. The proof is left as an exercise.

**Proposition 3.2.4.** *If  $p, q, u$  are bounded near  $x_0$ , then  $f \sim_u p$  and  $g \sim_u q$  at  $x_0$  imply  $fg \sim_u pq$  at  $x_0$ .*

The following is the general statement about composition of approximations.

**Proposition 3.2.5.** *Suppose  $f(x) \sim_{u(x)} p(x)$  at  $x_0$  and  $g(y) \sim_{v(y)} q(y)$  at  $y_0 = f(x_0)$ . Suppose*

1.  $p$  is continuous at  $x_0$ ,
2.  $u$  is bounded near  $x_0$ ,
3.  $v(f(x)) \leq Au(x)$  near  $x_0$  for a constant  $A$ ,
4.  $|q(y_1) - q(y_2)| \leq B|y_1 - y_2|$  for  $y_1, y_2$  near  $y_0$ .

*Then  $g(f(x)) \sim_{u(x)} q(p(x))$ .*

*Proof.* By  $f \sim_u p$  and  $g \sim_v q$ , for any  $\epsilon_1, \epsilon_2 > 0$ , there are  $\delta_1, \delta_2 > 0$ , such that

$$|\Delta x| = |x - x_0| < \delta_1 \implies |f(x) - p(x)| \leq \epsilon_1 u(x), \quad (3.2.12)$$

$$|\Delta y| = |y - y_0| < \delta_2 \implies |g(y) - q(y)| \leq \epsilon_2 v(y). \quad (3.2.13)$$

By Exercise 3.45, we have  $p(x_0) = f(x_0) = y_0$ . By the first, second and third conditions, for any  $\epsilon_3 > 0$ , we may further assume that

$$|\Delta x| = |x - x_0| < \delta_1 \implies |p(x) - y_0| \leq \epsilon_3, \quad u(x) \leq C, \quad v(f(x)) \leq Au(x). \quad (3.2.14)$$

By the fourth condition, we may further assume that

$$|y_1 - y_0| < \delta_2, \quad |y_2 - y_0| < \delta_2 \implies |q(y_1) - q(y_2)| \leq B|y_1 - y_2|. \quad (3.2.15)$$

Then  $|\Delta x| < \delta_1$  implies

$$|f(x) - y_0| \leq |f(x) - p(x)| + |p(x) - y_0| \leq \epsilon_1 u(x) + \epsilon_3 \leq C\epsilon_1 + \epsilon_3.$$

This means that, if we can arrange to have

$$C\epsilon_1 + \epsilon_3 < \delta_2, \quad (3.2.16)$$

then by (3.2.14),  $|\Delta x| < \delta_1$  implies that  $|f(x) - y_0| < \delta_2$  and  $|p(x) - y_0| \leq \epsilon_3 < \delta_2$ . Further by (3.2.15) and (3.2.12), we get

$$|q(f(x)) - q(p(x))| \leq B|f(x) - p(x)| \leq B\epsilon_1 u(x),$$

and by (3.2.13) and (3.2.14), we get

$$|g(f(x)) - q(f(x))| \leq \epsilon_2 v(f(x)) \leq \epsilon_2 Au(x).$$

Therefore  $|\Delta x| < \delta_1$  implies

$$\begin{aligned} |g(f(x)) - q(p(x))| &\leq |g(f(x)) - q(f(x))| + |q(f(x)) - q(p(x))| \\ &\leq \epsilon_2 Au(x) + B\epsilon_1 u(x) = (A\epsilon_2 + B\epsilon_1)u(x). \end{aligned}$$

The estimation suggests that, for any  $\epsilon > 0$ , we first choose  $\epsilon_2 = \frac{\epsilon}{2A}$ . Then for this  $\epsilon_2$ , we find  $\delta_2$  such that (3.2.13) and (3.2.15) hold. Then for this  $\delta_2$ , we take

$\epsilon_1 = \min \left\{ \frac{\delta_2}{2C}, \frac{\epsilon}{2B} \right\}$  and  $\epsilon_3 = \frac{\delta_2}{3}$ , so that (3.2.16) is satisfied, and  $A\epsilon_2 + B\epsilon_1 \leq \epsilon$ .

Next for  $\epsilon_1, \epsilon_3$ , we find  $\delta_1$ , such that (3.2.12) and (3.2.14) hold. After these choices, we may conclude that  $|\Delta x| < \delta_1$  implies

$$|g(f(x)) - q(p(x))| \leq (A\epsilon_2 + B\epsilon_1)u(x) \leq \epsilon u(x). \quad \square$$

**Example 3.2.1.** We explain the Leibniz rule by the general principle of approximation.

Suppose we have linear approximations  $f \sim_{|\Delta x|} p = a + b\Delta x$  and  $g \sim_{|\Delta x|} q = c + d\Delta x$ . Then by Proposition 3.2.4, we have

$$fg \sim_{|\Delta x|} pq = ac + (ad + bc)\Delta x + bd(\Delta x)^2.$$

By the first property in Exercise 3.3, we have  $ac + (ad + bc)\Delta x \sim_{|\Delta x|} ac + (ad + bc)\Delta x$ . By Exercise 3.47, we have  $bd(\Delta x)^2 \sim_{|\Delta x|} 0$ . Then by Proposition 3.2.3, we have

$$ac + (ad + bc)\Delta x + bd(\Delta x)^2 \sim_{|\Delta x|} ac + (ad + bc)\Delta x + 0 = ac + (ad + bc)\Delta x.$$

By the third property in Exercise 3.3, we conclude

$$fg \sim_{|\Delta x|} pq = ac + (ad + bc)\Delta x.$$

Exercise 3.43. Prove Proposition 3.2.4.

Exercise 3.44. Derive the chain rule for linear approximation from Proposition 3.2.5.

Exercise 3.45. Prove that  $f \sim_u p$  at  $x_0$  implies  $f(x_0) = p(x_0)$ .

Exercise 3.46. Prove that  $f \sim_u p$  if and only if  $f - p \sim_u 0$ .

Exercise 3.47. Suppose  $u(x) \neq 0$  for  $x \neq x_0$ . Prove that  $f \sim_u 0$  at  $x_0$  if and only if  $f(x_0) = 0$  and  $\lim_{x \rightarrow x_0} \frac{f(x)}{u(x)} = 0$ .

Exercise 3.48. Suppose  $\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = 0$ . Prove that  $f(x) + o(f(x)) = g(x) + o(g(x))$  implies  $f(x) = g(x) + o(g(x))$ .

### 3.3 Mean Value Theorem

The Mean Value Theorem says that, under good conditions, we have

$$\frac{f(b) - f(a)}{b - a} = f'(c) \text{ for some } c \in (a, b).$$

If  $f(x)$  is the distance one travels by the time  $x$ , then the left side is the average traveling speed from time  $a$  to time  $b$ . The Mean Value Theorem reflects the intuition that one must travel at exactly the average speed at some some moment during the trip. The theorem is used to establish many important results in differentiation theory. A proof of the theorem uses the extreme values of differentiable functions.

## Maximum and Minimum

A function  $f(x)$  has a *local maximum* at  $x_0$  if  $f(x_0)$  is the biggest value among all the values near  $x_0$ . In other words, there is  $\delta > 0$ , such that

$$|x - x_0| < \delta \implies f(x_0) \geq f(x).$$

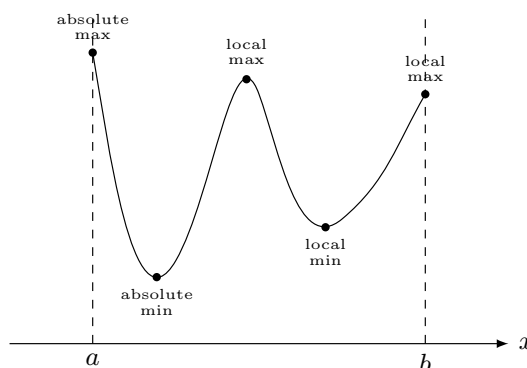
Similarly,  $f(x)$  has a *local minimum* at  $x_0$  if there is  $\delta > 0$ , such that

$$|x - x_0| < \delta \implies f(x_0) \leq f(x).$$

Local maxima and local minima are also called *local extrema*.

In the definition above, the function is assumed to be defined on both sides of  $x_0$ . Similar definition can be made when the function is defined on only one side of  $x_0$ . For example, a function  $f(x)$  defined on a bounded closed interval  $[a, b]$  has local maximum at  $a$  if there is  $\delta > 0$ , such that

$$0 \leq x - a < \delta \implies f(a) \geq f(x).$$



**Figure 3.3.1.** *maximum and minimum*

A function  $f(x)$  has an *absolute maximum* at  $x_0$  if  $f(x_0)$  is the biggest value among all the values of  $f(x)$ . In other words, we have  $f(x_0) \geq f(x)$  for any  $x$  in the domain of  $f$ . The *absolute minimum* is similarly defined. Absolute extrema are clearly also local extrema.

**Proposition 3.3.1.** *Suppose a function  $f(x)$  is defined on both sides of  $x_0$  and has a local extreme at  $x_0$ . If  $f(x)$  is differentiable at  $x_0$ , then  $f'(x_0) = 0$ .*

*Proof.* Suppose  $f(x)$  is differentiable at  $x_0$ , with  $f'(x_0) > 0$ . Fix any  $0 < \epsilon < f'(x_0)$ . Then there is  $\delta > 0$ , such that

$$\begin{aligned} |x - x_0| < \delta &\implies |f(x) - f(x_0) - f'(x_0)(x - x_0)| \leq \epsilon|x - x_0| \\ &\iff -\epsilon|x - x_0| \leq f(x) - f(x_0) - f'(x_0)(x - x_0) \leq \epsilon|x - x_0|. \end{aligned}$$

For  $x_0 < x < x_0 + \delta$ , we have  $x - x_0 = |x - x_0|$ , and this implies

$$f(x) - f(x_0) \geq f'(x_0)(x - x_0) - \epsilon|x - x_0| = (f'(x_0) - \epsilon)|x - x_0| > 0.$$

For  $x_0 - \delta < x < x_0$ , we have  $x - x_0 = -|x - x_0|$ , and this implies

$$f(x) - f(x_0) \leq f'(x_0)(x - x_0) + \epsilon|x - x_0| = -(f'(x_0) - \epsilon)|x - x_0| < 0.$$

Therefore  $f(x_0)$  is strictly smaller than the right side and strictly bigger than the left side. In particular,  $x_0$  is not a local extreme. The argument for the case  $f'(x_0) < 0$  is similar.  $\square$

We emphasize that the proof makes critical use of both sides of  $x_0$ . For a function  $f(x)$  defined on a bounded closed interval  $[a, b]$ , this means that the proposition may be applied to the *interior* points of the interval where the function is differentiable. Therefore a local extreme point  $x_0$  must be one of the following three cases.

1.  $a < x_0 < b$ ,  $f'(x_0)$  does not exist.
2.  $a < x_0 < b$ ,  $f'(x_0)$  exists and is 0.
3.  $x_0 = a$  or  $b$ .

Typically, the three possibilities would provide finitely many candidates for the potential local extrema. If we take the maximum and minimum of the values at these points, then we get the absolute maximum and the absolute minimum.

**Example 3.3.1.** If  $f(x)$  is differentiable at  $x_0$ , then  $f(x)^2$  is differentiable because this is the composition with a differentiable function  $y^2$ .

Conversely, suppose  $f(x)$  is continuous at  $x_0$ . If  $f(x)^2$  is differentiable at  $x_0$  and  $f(x_0) > 0$ , then we have  $f(x) = \sqrt{f(x)^2}$  near  $x_0$ . Since the square root function is differentiable at  $f(x_0)$ , the composition  $f(x)$  is also differentiable at  $x_0$ . Similarly, if  $f(x_0) < 0$ , then  $f(x) = -\sqrt{f(x)^2}$  near  $x_0$  and is differentiable at  $x_0$ .

So we conclude that, if  $f(x_0) \neq 0$ , then  $f(x)^2$  is differentiable at  $x_0$  if and only if  $f(x)$  is differentiable at  $x_0$ .

In case  $f(x_0) = 0$ ,  $x_0$  is a local minimum of  $f(x)^2$ . Therefore the derivative of  $f(x)^2$  at  $x_0$  vanishes, and the linear approximation of  $f(x)^2$  at  $x_0$  is  $0 + 0\Delta x = 0$ . This means  $f(x)^2 = o(x - x_0)$ , which is equivalent to  $f(x) = o(\sqrt{|x - x_0|})$ . We conclude that, if  $f(x_0) = 0$ , then  $f(x)^2$  is differentiable at  $x_0$  if and only if  $f(x) = o(\sqrt{|x - x_0|})$ .

**Exercise 3.49.** Suppose  $f(x)$  is continuous at  $x_0$ . Find the conditions for  $f(x)^3$ ,  $f(x)^{\frac{1}{3}}$ ,  $f(x)^{\frac{2}{3}}$  to be differentiable at  $x_0$ . What about  $f(x)^{\frac{m}{n}}$  in general, where  $m, n$  are natural numbers?

**Exercise 3.50.** Suppose  $f(x)$  is continuous at 0.

1. Prove that if  $f(0) \notin \frac{1}{2}\pi + \mathbb{Z}\pi$ , then  $\sin f(x)$  is differentiable at 0 if and only if  $f(x)$  is differentiable at 0.



2. Prove that if  $f(0) = \frac{1}{2}\pi$ , then  $\sin f(x)$  is differentiable at 0 if and only if  $f(x) = \frac{1}{2}\pi + o(\sqrt{|x|})$  near 0.

**Exercise 3.51.** If  $f(x)$  has local maximum at  $x_0$  and is right differentiable at  $x_0$ , prove that  $f'_+(x_0) \leq 0$ . State the similar results for other local extrema and one sided derivatives.

**Exercise 3.52.** Suppose  $f(x)$  is differentiable on  $[a, b]$  ( $f$  is differentiable on  $(a, b)$ , and  $f'_+(a), f'_-(b)$  exist at the two ends). If  $f'_+(a) < 0$  and  $f'_-(b) > 0$ , prove that the absolute minimum of  $f$  lies in  $(a, b)$ .

**Exercise 3.53 (Darboux's Intermediate Value Theorem).** Suppose  $f(x)$  is differentiable on  $[a, b]$ , and  $\gamma$  lies between  $f'_+(a)$  and  $f'_-(b)$ . Prove that  $f(x) - \gamma x$  has an absolute extreme in  $(a, b)$ . Then prove that  $\gamma$  is the value of  $f$  somewhere in  $(a, b)$ .

**Exercise 3.54.** Find a function  $f(x)$  that is differentiable everywhere on  $[-1, 1]$ , yet  $f'(x)$  is not continuous. The examples shows that Darboux's Intermediate Value Theorem is not a consequence of the usual Intermediate Value Theorem.

## Mean Value Theorem

**Theorem 3.3.2 (Mean Value Theorem).** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Then there is  $c \in (a, b)$ , such that

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

Geometrically, the quotient on the left is the slope of the line segment that connects the end points of the graph of  $f(x)$  on  $[a, b]$ . The theorem says that the line segment is parallel to some tangent line.

The conclusion of the theorem can also be written as

$$f(b) - f(a) = f'(c)(b - a) \text{ for some } c \in (a, b),$$

or

$$f(x + \Delta x) - f(x) = f'(x + \theta \Delta x) \Delta x \text{ for some } 0 < \theta < 1.$$

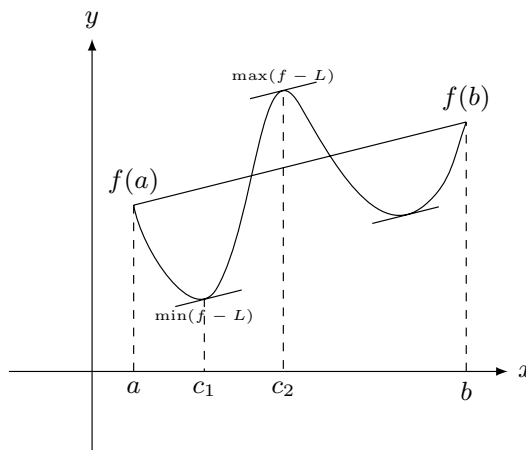
Also note that  $a$  and  $b$  may be exchanged in the theorem, so that we do not have to insist  $a < b$  (or  $\Delta x > 0$ ) for the equalities to hold.

*Proof.* The line connecting  $(a, f(a))$  and  $(b, f(b))$  is

$$L(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

As suggested by Figure 3.3.2, the tangent lines parallel to  $L(x)$  can be found where the difference

$$h(x) = f(x) - L(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)$$



**Figure 3.3.2.** *Mean Value Theorem.*

reaches maximum or minimum.

Since  $h(x)$  is continuous, by Theorem 2.4.2, it reaches maximum and minimum at  $c_1, c_2 \in [a, b]$ . If both  $c_1$  and  $c_2$  are end points  $a$  and  $b$ , then the maximum and the minimum of  $h(x)$  are  $h(a) = h(b) = 0$ . This implies  $h(x)$  is constantly zero on the interval, so that  $h'(c) = 0$  for any  $c \in [a, b]$ . If one of  $c_1$  and  $c_2$ , denoted  $c$ , is not an end point, then by Proposition 3.3.1, we have  $h'(c) = 0$ . In any case, we have  $c \in (a, b)$  satisfying

$$h'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0. \quad \square$$

**Exercise 3.55.** Find  $c$  in the Mean Value Theorem.

1.  $x^3$  on  $[-1, 1]$ .
2.  $\frac{1}{x}$  on  $[1, 2]$ .
3.  $2^x$  on  $[0, 1]$ .

**Exercise 3.56.** Let  $f(x) = |x - 1|$ . Is there  $c \in [0, 3]$  such that  $f(3) - f(0) = f'(c)(3 - 0)$ ? Does your conclusion contradict the Mean Value Theorem?

**Exercise 3.57.** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Prove that  $f(x)$  is Lipschitz (see Exercise 2.42) if and only if  $f'(x)$  is bounded on  $(a, b)$ .

**Exercise 3.58.** Determine the uniform continuity on  $(0, 1]$ ,  $[1, +\infty)$ ,  $(0, +\infty)$ .

1.  $x^p$ .
2.  $\sin x^p$ .
3.  $x^p \sin \frac{1}{x}$ .

**Exercise 3.59 (Rolle<sup>12</sup>'s Theorem).** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Prove that if  $f(a) = f(b)$ , then  $f'(c) = 0$  for some  $c \in (a, b)$ .

<sup>12</sup>Michel Rolle, born 1652 in Ambert (France), died 1719 in Paris (France). Rolle invented the notion  $\sqrt[n]{x}$  for the  $n$ -th root of  $x$ . His theorem appeared in an obscure book in 1691.

**Exercise 3.60.** Suppose  $f(x)$  is continuous on  $[a, b]$  and is left and right differentiable on  $(a, b)$ . Prove that there is  $c \in (a, b)$ , such that  $\frac{f(b) - f(a)}{b - a}$  lies between  $f'_-(c)$  and  $f'_+(c)$ .

**Exercise 3.61.** Suppose  $f(x)$  is continuous at  $x_0$  and differentiable on  $(x_0 - \delta, x_0) \cup (x_0, x_0 + \delta)$ . Prove that if  $\lim_{x \rightarrow x_0} f'(x) = l$  converges, then  $f(x)$  is differentiable at  $x_0$  and  $f'(x_0) = l$ .

**Exercise 3.62.** Suppose  $f(x)$  has continuous derivative on a bounded closed interval  $[a, b]$ . Prove that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|\Delta x| < \delta \implies |f(x + \Delta x) - f(x) - f'(x)\Delta x| \leq \epsilon |\Delta x|.$$

In other words,  $f(x)$  is *uniformly differentiable*.

**Exercise 3.63.** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Suppose  $f(a) = 0$  and  $|f'(x)| \leq A|f(x)|$  for a constant  $A$ .

1. Prove that  $f(x) = 0$  on  $\left[a, a + \frac{1}{2A}\right]$ .
2. Prove that  $f(x) = 0$  on the whole interval  $[a, b]$ .

**Exercise 3.64.** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Suppose  $|\lambda(y)| \leq A|y|$  for a constant  $A$  and sufficiently small  $y$ . Suppose  $f(a) = 0$  and  $|f'(x)| \leq |\lambda(f(x))|$ .

1. Prove that  $f(x) = 0$  on  $[a, a + \delta]$  for some  $\delta > 0$ .
2. Prove that  $f(x) = 0$  on the whole interval  $[a, b]$ .

Note that if  $\lambda$  is differentiable at 0 and  $\lambda(0) = 0$ , then  $\lambda$  has the required property.

## Cauchy's Mean Value Theorem

A very useful extension of the Mean Value Theorem is due to Cauchy.

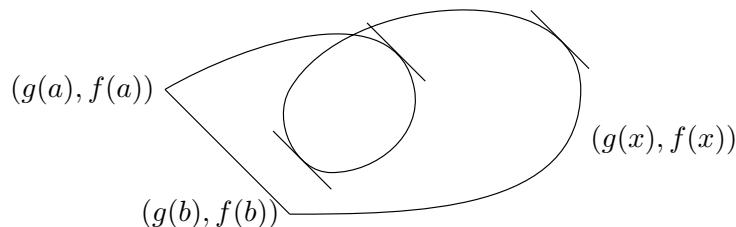
**Theorem 3.3.3 (Cauchy's Mean Value Theorem).** Suppose  $f(x)$  and  $g(x)$  are continuous on  $[a, b]$  and differentiable on  $(a, b)$ . If  $g'(x)$  is never zero, then there is  $c \in (a, b)$ , such that

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}.$$

Geometrically, consider  $(g(x), f(x))$  as a parameterized curve in  $\mathbb{R}^2$ . The vector from one point at  $x = a$  to another point at  $x = b$  is  $(g(b) - g(a), f(b) - f(a))$ . Cauchy's Mean Value Theorem says that it should be parallel to a tangent vector  $(g'(c), f'(c))$  at a point  $(g(c), f(c))$  on the curve. This suggests that the theorem may be proved by imitating the proof of the Mean Value Theorem, by considering

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}(g(x) - g(a)).$$

The details are left to the reader.



**Figure 3.3.3.** *Cauchy's Mean Value Theorem.*

**Exercise 3.65.** Find  $c$  in Cauchy's Mean Value Theorem.

1.  $f(x) = x^3$ ,  $g(x) = x^2$  on  $[1, 2]$ .
2.  $f(x) = \sin x$ ,  $g(x) = \cos x$  on  $\left[0, \frac{\pi}{2}\right]$ .
3.  $f(x) = e^{2x}$ ,  $g(x) = e^x$  on  $[-a, a]$ .

**Exercise 3.66.** Explain why  $g(a) \neq g(b)$  under the assumption of Cauchy's Mean Value Theorem.

**Exercise 3.67.** Prove Cauchy's Mean Value Theorem.

**Exercise 3.68.** Suppose the assumption that  $g'(x)$  is never zero is replaced by the (weaker) assumption that  $g(x)$  is strictly monotone. What can you say about the conclusion of Cauchy's Mean Value Theorem.

## Constant Function

A consequence of the Mean Value Theorem is that a non-changing quantity must be a constant.

**Proposition 3.3.4.** *Suppose  $f'(x) = 0$  for all  $x$  on an interval. Then  $f(x)$  is a constant on the interval.*

For any two points  $x_1$  and  $x_2$  in the interval, the Mean Value Theorem gives

$$f(x_1) - f(x_2) = f'(c)(x_1 - x_2) = 0(x_1 - x_2) = 0,$$

for some  $c$  between  $x_1$  and  $x_2$ . This proves the proposition.

**Example 3.3.2.** Suppose  $f(x)$  satisfies  $f'(x) = af(x)$ , where  $a$  is a constant. Then

$$(e^{-ax}f(x))' = -ae^{-ax}f(x) + e^{-ax}f'(x) = -ae^{-ax}f(x) + ae^{-ax}f(x) = 0.$$

Therefore  $e^{-ax}f(x) = c$  is a constant, and  $f(x) = ce^{ax}$  is a scalar multiple of the exponential function.

**Exercise 3.69.** Suppose  $f(x)$  satisfies  $f'(x) = xf(x)$ . Prove that  $f(x) = ce^{\frac{x^2}{2}}$  for some constant  $c$ .

**Exercise 3.70.** Suppose  $f(x)$  satisfies  $|f(x) - f(y)| \leq |x - y|^p$  for some constant  $p > 1$ . Prove that  $f(x)$  is constant.

**Exercise 3.71.** Prove that if  $f'(x) = g'(x)$  for all  $x$ , then  $f(x) = g(x) + c$  for some constant  $c$ .

## Monotone Function

To find out whether a function  $f(x)$  is increasing near  $x_0$ , we note that the linear approximation  $f(x_0) + f'(x_0)\Delta x$  is increasing if and only if the coefficient  $f'(x_0) \geq 0$ . Because of the approximation, we may expect that the condition  $f'(x_0) \geq 0$  implies that  $f(x)$  is also increasing near  $x_0$ . In fact, Exercise 3.78 shows that such an expectation is wrong. Still, the idea inspires the following correct statement.

**Proposition 3.3.5.** *Suppose  $f(x)$  is continuous on an interval and is differentiable on the interior of the interval. Then  $f(x)$  is increasing if and only if  $f'(x) \geq 0$ . Moreover, if  $f'(x) > 0$ , then  $f(x)$  is strictly increasing.*

We have similar result for decreasing functions.

Combined with Theorem 2.5.3, we find that  $f'(x) \neq 0$  near  $x_0$  implies the invertibility of  $f(x)$  near  $x_0$ . However, Exercise 3.41 shows that just  $f'(x_0) \neq 0$  alone does not imply the invertibility.

*Proof.* Suppose  $f(x)$  is increasing. Then either  $f(y) = f(x)$ , or  $f(y) - f(x)$  has the same sign as  $y - x$ . Therefore  $\frac{f(y) - f(x)}{y - x} \geq 0$  for any  $x \neq y$ . By taking  $y \rightarrow x$ , we get  $f'(x) \geq 0$  for any  $x$ .

Conversely, for  $x < y$ , we have  $f(y) - f(x) = f'(c)(y - x)$  for some  $c \in (x, y)$  by the Mean Value Theorem. Then the assumption  $f'(c) \geq 0$  implies  $f(y) - f(x) \geq 0$ , and the assumption  $f'(c) > 0$  implies  $f(y) - f(x) > 0$ .  $\square$

**Exercise 3.72.** Suppose  $f(x)$  and  $g(x)$  are continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Prove that if  $f(a) \leq g(a)$  and  $f'(x) \leq g'(x)$  for  $x \in (a, b)$ , then  $f(x) \leq g(x)$  on  $[a, b]$ . Moreover, if one of the inequalities in the assumption is strict, then  $f(x) < g(x)$  on  $[a, b]$ . Finally, state the similar result for an interval on the left of  $a$ .

**Exercise 3.73.** Suppose  $f(x)$  is left and right differentiable on an interval. Prove that if  $f'_+(x) \geq 0$  and  $f'_-(x) \geq 0$ , then  $f(x)$  is increasing. Moreover, if the inequalities are strict, then  $f(x)$  is strictly increasing.

**Exercise 3.74 (Young<sup>13</sup>'s Inequality).** Suppose  $p, q$  are real numbers satisfying  $\frac{1}{p} + \frac{1}{q} = 1$ .

<sup>13</sup>William Henry Young, born 1863 in London (England), died 1942 in Lausanne (Switzerland).

1. For  $x > 0$ , prove that

$$x^{\frac{1}{p}} \leq \frac{1}{p}x + \frac{1}{q} \text{ for } p > 1, \quad x^{\frac{1}{p}} \geq \frac{1}{p}x + \frac{1}{q} \text{ for } p < 1.$$

2. For  $x, y > 0$ , prove that

$$xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q \text{ for } p > 1, \quad xy \geq \frac{1}{p}x^p + \frac{1}{q}y^q \text{ for } p < 1.$$

When does the equality hold?

**Exercise 3.75 (Hölder<sup>14</sup>'s Inequality).** Suppose  $p, q > 0$  satisfy  $\frac{1}{p} + \frac{1}{q} = 1$ . For positive numbers  $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n$ , by taking  $a = \frac{a_i}{(\sum a_i^p)^{\frac{1}{p}}}$  and  $b = \frac{b_i}{(\sum b_i^q)^{\frac{1}{q}}}$  in Young's inequality, prove that

$$\sum a_i b_i \leq \left( \sum a_i^p \right)^{\frac{1}{p}} \left( \sum b_i^q \right)^{\frac{1}{q}}.$$

When does the equality hold?

**Exercise 3.76 (Minkowski<sup>15</sup>'s Inequality).** Suppose  $p > 1$ . By applying Hölder's inequality to  $a_1, a_2, \dots, a_n, (a_1 + b_1)^{p-1}, (a_2 + b_2)^{p-1}, \dots, (a_n + b_n)^{p-1}$  and then to  $b_1, b_2, \dots, b_n, (a_1 + b_1)^{p-1}, (a_2 + b_2)^{p-1}, \dots, (a_n + b_n)^{p-1}$ , prove that

$$\left( \sum (a_i + b_i)^p \right)^{\frac{1}{p}} \leq \left( \sum a_i^p \right)^{\frac{1}{p}} + \left( \sum b_i^p \right)^{\frac{1}{p}}.$$

When does the equality hold?

**Exercise 3.77.** Suppose  $f(x)$  is continuous for  $x \geq 0$  and differentiable for  $x > 0$ . Prove that if  $f'(x)$  is strictly increasing and  $f(0) = 0$ , then  $\frac{f(x)}{x}$  is also strictly increasing.

**Exercise 3.78.** Suppose  $p > 1$ ,  $f(x) = x + |x|^p \sin \frac{1}{x}$  for  $x \neq 0$ , and  $f(0) = 0$ .

1. Show that  $f(x)$  is differentiable everywhere, with  $f'(0) = 1$ .
2. Show that if  $p < 2$ , then  $f(x)$  is not monotone on  $(0, \delta)$  for any  $\delta > 0$ .

## L'Hôpital's Rule

The derivative can be used to compute the function limits of the types  $\frac{0}{0}$  or  $\frac{\infty}{\infty}$ .

Young discovered a form of Lebesgue integration independently. He wrote an influential book "The fundamental theorems of the differential calculus" in 1910.

<sup>14</sup>Otto Ludwig Hölder, born 1859 in Stuttgart (Germany), died 1937 in Leipzig (Germany). He discovered the inequality in 1884. Hölder also made fundamental contributions to the group theory.

<sup>15</sup>Hermann Minkowski, born 1864 in Alexotas (Russia, now Kaunas of Lithuania), died 1909 in Göttingen (Germany). Minkowski's fundamental contribution to geometry provided the mathematical foundation of Einstein's theory of relativity.

**Proposition 3.3.6** (L'Hôpital<sup>16</sup>'s Rule). *Suppose  $f(x)$  and  $g(x)$  are differentiable on  $(a - \delta, a) \cup (a, a + \delta)$  for some  $\delta > 0$ . Assume*

1. *Either  $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = 0$  or  $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = \infty$ .*
2.  *$\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$  converges.*

*Then  $\lim_{x \rightarrow a} \frac{f(x)}{g(x)}$  also converges and  $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$ .*

One should not blindly use l'Hôpital's rule whenever it comes to the limit of a quotient, for three reasons.

First, one always needs to make sure that both conditions are satisfied. The following counterexample

$$\lim_{x \rightarrow 0} \frac{1+x}{2+x} = \frac{1}{2}, \quad \lim_{x \rightarrow 0} \frac{(1+x)'}{(2+x)'} = \lim_{x \rightarrow 0} \frac{1}{1} = 1.$$

shows the necessity of the first condition. The second condition means that the convergence of  $\frac{f'}{g'}$  implies the convergence of  $\frac{f}{g}$ . The converse of this implication is not necessarily true.

Second, using l'Hôpital's rule may lead to logical fallacy. For example, the following application of l'Hôpital's rule is mathematically correct in the sense that both conditions are satisfied

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{(\sin x)'}{x'} = \lim_{x \rightarrow 0} \frac{\cos x}{1} = \cos 0 = 1.$$

Logically, however, the argument above is circular: The conclusion  $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$  means that the derivative of  $\sin x$  at 0 is 1. Yet the derivative of  $\sin x$  is used in the argument.

Third, using l'Hôpital's rule is often more complicated than the use of approximations. For fairly complicated functions, it is simply complicated to compute the derivatives.

*Proof of l'Hôpital's Rule.* We will prove for the limit of the type  $\lim_{x \rightarrow a+}$  only, with  $a$  a finite number. Thus  $f(x)$  and  $g(x)$  are assumed to be differentiable on  $(a, a + \delta)$ .

First assume  $\lim_{x \rightarrow a+} f(x) = \lim_{x \rightarrow a+} g(x) = 0$ . Then  $f(x)$  and  $g(x)$  can be extended to continuous functions on  $[a, a + \delta)$  by defining  $f(a) = g(a) = 0$ . Cauchy's Mean Value Theorem then tells us that for any  $x \in [a, a + \delta)$ , we have

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)} \quad (3.3.1)$$

<sup>16</sup>Guillaume Francois Antoine Marquis de l'Hôpital, born 1661 in Paris (France), died 1704 in Paris (France). His famous rule was found in his 1696 book "Analyse des infiniment petits pour l'intelligence des lignes courbes", which was the first textbook in differential calculus.

for some  $c \in (a, x)$  (and  $c$  depends on  $x$ ). As  $x \rightarrow a^+$ , we have  $c \rightarrow a^+$ . Therefore if the limit on the right of (3.3.1) converges, so does the limit on the left, and the two limits are the same.

Now consider the case  $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = \infty$ . The technical difficulty here is that the functions cannot be extended to  $x = a$  as before. Still, we try to establish something similar to (3.3.1) by replacing  $\frac{f(x) - f(a)}{g(x) - g(a)}$  with  $\frac{f(x) - f(b)}{g(x) - g(b)}$ , where  $b > a$  is very close to  $a$ . The second equality in (3.3.1) still holds. Although the first equality no longer holds, it is sufficient to show that  $\frac{f(x)}{g(x)}$  and  $\frac{f(x) - f(b)}{g(x) - g(b)}$  are very close. Of course all these should be put together in logical order.

Denote  $\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)} = l$ . For any  $\epsilon > 0$ , there is  $\delta' > 0$ , such that

$$a < x < b = a + \delta' \implies \left| \frac{f'(x)}{g'(x)} - l \right| < \epsilon.$$

Then by Cauchy's Mean Value Theorem,

$$a < x < b \implies \left| \frac{f(x) - f(b)}{g(x) - g(b)} - l \right| = \left| \frac{f'(c)}{g'(c)} - l \right| < \epsilon, \quad (3.3.2)$$

where  $a < x < c < b$ . In particular, the quotient  $\frac{f(x) - f(b)}{g(x) - g(b)}$  is bounded. Moreover, since  $b$  has been fixed, by the assumption  $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = \infty$ , we have

$$\lim_{x \rightarrow a^+} \frac{1 - \frac{g(b)}{g(x)}}{1 - \frac{f(b)}{f(x)}} = 1.$$

Therefore, there is  $\delta' \geq \delta > 0$ , such that

$$\begin{aligned} a < x < a + \delta &\implies \left| \frac{f(x)}{g(x)} - \frac{f(x) - f(b)}{g(x) - g(b)} \right| \\ &= \left| \frac{f(x) - f(b)}{g(x) - g(b)} \right| \left| \frac{1 - \frac{g(b)}{g(x)}}{1 - \frac{f(b)}{f(x)}} - 1 \right| < \epsilon. \end{aligned}$$

Since  $a < x < a + \delta$  implies  $a < x < b$ , the conclusion of (3.3.2) also holds. Thus

$$\begin{aligned} a < x < a + \delta &\implies \left| \frac{f(x)}{g(x)} - l \right| \\ &\leq \left| \frac{f(x)}{g(x)} - \frac{f(x) - f(b)}{g(x) - g(b)} \right| + \left| \frac{f(x) - f(b)}{g(x) - g(b)} - l \right| < 2\epsilon. \quad \square \end{aligned}$$

**Exercise 3.79.** Discuss whether l'Hôpital's rule can be applied to the limits.



$$1. \lim_{x \rightarrow \infty} \frac{x + \sin x}{x - \cos x}, \quad 2. \lim_{x \rightarrow +\infty} \frac{x}{x + \sin x}, \quad 3. \lim_{x \rightarrow 0} \frac{x^2 \sin \frac{1}{x}}{\sin x}.$$

Exercise 3.80. The Mean Value Theorem tells us

$$\log(1+x) - \log 1 = x \frac{1}{1+\theta x}, \quad e^x - 1 = x e^{\theta x},$$

for some  $0 < \theta < 1$ . Prove that in both cases,  $\lim_{x \rightarrow 0} \theta = \frac{1}{2}$ .

Exercise 3.81. Use l'Hôpital's rule to calculate the limit in Example 3.4.7. What do you observe?

Exercise 3.82. Prove l'Hôpital's rule for the case  $a = +\infty$ . Moreover, discuss l'Hôpital's rule for the case  $l = \infty$ .

## 3.4 High Order Approximation

**Definition 3.4.1.** A function  $f(x)$  is  $n$ -th order differentiable at  $x_0$  if it is approximated by a polynomial of degree  $n$

$$p(x) = a_0 + a_1 \Delta x + a_2 \Delta x^2 + \cdots + a_n \Delta x^n, \quad \Delta x = x - x_0.$$

The  $n$ -th order approximation means that, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|\Delta x| < \delta \implies |f(x) - p(x)| \leq \epsilon |\Delta x|^n.$$

The definition means  $f \sim_{|\Delta x|^n} p$ , or

$$f(x) = a_0 + a_1 \Delta x + a_2 \Delta x^2 + \cdots + a_n \Delta x^n + o(\Delta x^n),$$

where  $o(\Delta x^n)$  denotes a function  $R(x)$  satisfying  $\lim_{\Delta x \rightarrow 0} \frac{R(x)}{\Delta x^n} = 0$ .

**Example 3.4.1.** By rewriting the function  $x^4$  as a polynomial in  $(x-1)$

$$f(x) = x^4 = (1 + (x-1))^4 = 1 + 4(x-1) + 6(x-1)^2 + 4(x-1)^3 + (x-1)^4,$$

we get high order approximations of  $x^4$  at 1

$$\begin{aligned} T_1(x) &= 1 + 4(x-1), \\ T_2(x) &= 1 + 4(x-1) + 6(x-1)^2, \\ T_3(x) &= 1 + 4(x-1) + 6(x-1)^2 + 4(x-1)^3, \\ T_n(x) &= 1 + 4(x-1) + 6(x-1)^2 + 4(x-1)^3 + (x-1)^4, \text{ for } n \geq 4. \end{aligned}$$

Therefore  $x^4$  is differentiable of any order. In general, a polynomial is differentiable of any order.

**Example 3.4.2.** Suppose we have a fifth order approximation

$$f(x) \sim_{|\Delta x|^5} p(x) = a_0 + a_1\Delta x + a_2\Delta x^2 + a_3\Delta x^3 + a_4\Delta x^4 + a_5\Delta x^5.$$

By the first part of Exercise 3.3 and Exercise 3.47, we get

$$\begin{aligned} a_0 + a_1\Delta x + a_2\Delta x^2 + a_3\Delta x^3 &\sim_{|\Delta x|^3} a_0 + a_1\Delta x + a_2\Delta x^2 + a_3\Delta x^3, \\ a_4\Delta x^4 + a_5\Delta x^5 &\sim_{|\Delta x|^3} 0. \end{aligned}$$

By Proposition 3.2.3, we may add the approximations together to get

$$p(x) \sim_{|\Delta x|^3} a_0 + a_1\Delta x + a_2\Delta x^2 + a_3\Delta x^3.$$

On the other hand, by Exercise 3.4,  $f(x) \sim_{|\Delta x|^5} p(x)$  implies  $f(x) \sim_{|\Delta x|^3} p(x)$ . Then by the third part of Exercise 3.3, we have

$$f(x) \sim_{|\Delta x|^3} a_0 + a_1\Delta x + a_2\Delta x^2 + a_3\Delta x^3.$$

In general, suppose two natural numbers satisfy  $m < n$ . If  $f(x)$  is  $n$ -th order differentiable at  $x_0$  with  $n$ -th order approximation  $p(x)$ , then  $f(x)$  is  $m$ -th order differentiable at  $x_0$  with the  $m$ -th order truncation of  $p(x)$  as the  $m$ -th order approximation.

**Example 3.4.3.** The quadratic (second order) approximation means that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|\Delta x| < \delta \implies |f(x) - a_0 - a_1\Delta x - a_2\Delta x^2| \leq \epsilon|\Delta x|^2.$$

By Example, 3.4.2, we know  $a_0 + a_1\Delta x$  is the linear approximation of  $f(x)$  at  $x_0$ . This means  $f(x)$  is continuous at  $x_0$ ,  $a_0 = f(x_0)$ , and  $a_1 = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$  converges. After getting  $a_0, a_1$ , the implication above means

$$0 < |\Delta x| < \delta \implies \left| \frac{f(x) - f(x_0) - f'(x_0)\Delta x}{\Delta x^2} - a_2 \right| \leq \epsilon.$$

This means exactly the convergence of  $a_2 = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - f'(x_0)\Delta x}{\Delta x^2}$ .

A consequence of the example is the uniqueness of quadratic approximation. In general, the  $n$ -th order approximation is a unique  $n$ -th order polynomial.

**Exercise 3.83.** Prove that if  $p$  and  $q$  are  $n$ -th order approximations of  $f$  and  $g$ , then  $p + q$  is the  $n$ -th order approximations of  $f + g$ .

**Exercise 3.84.** Prove that if  $p$  and  $q$  are  $n$ -th order approximations of  $f$  and  $g$ , then the  $n$ -th order truncation of  $pq$  is the  $n$ -th order approximations of  $fg$ .

**Exercise 3.85.** Prove that if  $p$  and  $q$  are  $n$ -th order approximations of  $f$  and  $g$ , then the  $n$ -th order truncation of  $q \circ p$  is the  $n$ -th order approximations of  $g \circ f$ .

**Exercise 3.86.** How are the high order approximations of  $f(x)$  and  $f(x^2)$  at 0 related?

Exercise 3.87. Directly prove the uniqueness of high order approximation: If

$$f(x) \sim_{|\Delta x|^n} p(x) = a_0 + a_1 \Delta x + a_2 \Delta x^2 + \cdots + a_n \Delta x^n,$$

$$f(x) \sim_{|\Delta x|^n} q(x) = b_0 + b_1 \Delta x + b_2 \Delta x^2 + \cdots + b_n \Delta x^n,$$

then  $p(x) = q(x)$  (i.e.,  $a_0 = b_0$ ,  $a_1 = b_1$ ,  $\dots$ ,  $a_n = b_n$ ).

Exercise 3.88. Prove that  $f(x)$  is  $n$  order differentiable at  $x_0$  if and only if the following are satisfied.

- $f$  is continuous at  $x_0$ . Let  $a_0 = f(x_0)$ .
- The limit  $a_1 = \lim_{x \rightarrow x_0} \frac{1}{\Delta x} (f(x) - a_0)$  converges.
- The limit  $a_2 = \lim_{x \rightarrow x_0} \frac{1}{\Delta x^2} (f(x) - a_0 - a_1 \Delta x)$  converges.
- $\dots$ .
- The limit  $a_n = \lim_{x \rightarrow x_0} \frac{1}{(\Delta x)^n} (f(x) - a_0 - a_1 \Delta x - \cdots - a_{n-1} \Delta x^{n-1})$  converges.

Moreover, the  $n$ -th order approximation is given by  $a_0 + a_1 \Delta x + a_2 \Delta x^2 + \cdots + a_n \Delta x^n$ .

Exercise 3.89. Suppose  $f(x)$  has  $n$ -th order approximation at  $x_0$  by polynomial  $p(x)$  of degree  $n$ . Prove that  $f$  is  $(n+k)$ -th order differentiable at  $x_0$  if and only if  $f(x) = p(x) + g(x) \Delta x^n$  for a function  $g(x)$  that is  $k$ -th order differentiable at  $x_0$ . The exercise is the high order generalization of Exercise 3.17. The high derivative version is Exercise 3.127.

Exercise 3.90. Define high order left and right differentiability. What is the relation between the usual high order differentiability and one sided high order differentiability?

Exercise 3.91. Prove that the high order approximation of an even function at 0 contains only terms of even power. What about an odd function?

Exercise 3.92. Suppose  $f(x)$  has cubic approximation  $p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$  at 0. Find the exact condition on  $p(x)$  such that  $f(\sqrt{|x|})$  is differentiable at 0.

Exercise 3.93. Suppose  $\lim_{x \rightarrow x_0} f(x) = 0$  and  $g(x) = f(x) + o(f(x)^2)$ .

1. Prove that  $\lim_{x \rightarrow x_0} g(x) = 0$  and  $f(x) = g(x) + o(g(x)^2)$ .
2. Prove that  $f(x)$  is second order differentiable at  $x_0$  if and only if  $g(x)$  is second order differentiable at  $x_0$ .

## Taylor Expansion

A function  $f(x)$  is differentiable on an open interval if it is differentiable at every point of the interval. Then the derivative  $f'(x)$  is also a function on the interval. If the function  $f'(x)$  is also differentiable on the interval, then we have the second order derivative function  $f''(x)$ . If the function  $f''(x)$  is again differentiable, then we have the third order derivative  $f'''(x)$ . In general, the  $n$ -th order derivative of  $f(x)$  is denoted  $f^{(n)}(x)$  or  $\frac{d^n f}{dx^n}$ . A function is *smooth* if it has derivatives of any order.

When we say  $f$  has  $n$ -th order derivative at  $x_0$ , we mean  $f$  has  $(n-1)$ st order derivative  $f^{(n-1)}$  on an interval containing  $x_0$ , and the function  $f^{(n-1)}$  is differentiable at  $x_0$ . By Proposition 3.1.5, the differentiability is equivalent to the existence of the  $n$ -th order derivative  $f^{(n)}(x_0) = (f^{(n-1)})'(x_0)$ .

**Theorem 3.4.2.** *Suppose  $f(x)$  has the  $n$ -th order derivative  $f^{(n)}(x_0)$  at  $x_0$ . Then*

$$T_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

*is an  $n$ -th order approximation of  $f(x)$  at  $x_0$ . In particular, the function is  $n$ -th order differentiable at  $x_0$ .*

The polynomial  $T_n(x)$  is the  $n$ -th order Taylor expansion. The theorem says that the existence of  $n$ -th order derivative at  $x_0$  implies the  $n$ -th order differentiability at  $x_0$ . Example 3.4.8 shows that the converse is not true.

One may also introduce  $n$ -th order one sided derivatives. See Exercise 3.100.

It is easy to compute the following high order derivatives

$$\begin{aligned}(x^p)^{(n)} &= p(p-1) \cdots (p-n+1)x^{p-n}, \\ ((ax+b)^p)^{(n)} &= p(p-1) \cdots (p-n+1)a^n(ax+b)^{p-n}, \\ (e^x)^{(n)} &= e^x, \\ (a^x)^{(n)} &= a^x(\log a)^n, \\ (\log|x|)^{(n)} &= (-1)^{n-1}(n-1)!x^{-n}.\end{aligned}$$

It is also easy to calculate the high order derivatives of sine and cosine functions. Then we get the following Taylor expansions at 0

$$\begin{aligned}(1+x)^p &= 1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3 \\ &\quad + \cdots + \frac{p(p-1) \cdots (p-n+1)}{n!}x^n + o(x^n), \\ \frac{1}{1-x} &= 1 + x + x^2 + x^3 + \cdots + x^n + o(x^n), \\ e^x &= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + o(x^n), \\ \log(1+x) &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots + \frac{(-1)^{n-1}}{n}x^n + o(x^n), \\ \sin x &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \cdots + \frac{(-1)^{n+1}}{(2n-1)!}x^{2n-1} + o(x^{2n}), \\ \cos x &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \cdots + \frac{(-1)^n}{(2n)!}x^{2n} + o(x^{2n+1}).\end{aligned}$$

Theorem 3.4.2 concludes that the remainder  $R_n(x) = f(x) - T_n(x)$  satisfies  $\lim_{\Delta x \rightarrow 0} \frac{R_n(x)}{\Delta x^n} = 0$ . Under slightly stronger assumption, more can be said about the remainder.

**Proposition 3.4.3** (Lagrange<sup>17</sup>). Suppose  $f(t)$  has  $(n+1)$ -st order derivative between  $x$  and  $x_0$ . Then there is  $c$  between  $x$  and  $x_0$ , such that

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}. \quad (3.4.1)$$

When  $n = 0$ , the conclusion is exactly the Mean Value Theorem. The formula (3.4.1) is called the *Lagrange form* of the remainder. Exercises 3.133 and 4.49 give two other formulae for the remainder.

*Proof.* We note that the remainder satisfies

$$R_n(x_0) = R'_n(x_0) = R''_n(x_0) = \cdots = R_n^{(n)}(x_0) = 0.$$

Therefore by Cauchy's Mean Value Theorem, we have

$$\begin{aligned} \frac{R_n(x)}{(x-x_0)^n} &= \frac{R_n(x) - R_n(x_0)}{(x-x_0)^n - (x_0-x_0)^n} = \frac{R'_n(c_1)}{n(c_1-x_0)^{n-1}} \\ &= \frac{R'_n(c_1) - R'_n(x_0)}{n((c_1-x_0)^{n-1} - (x_0-x_0)^{n-1})} = \frac{R''_n(c_2)}{n(n-1)(c_2-x_0)^{n-2}} \\ &= \cdots = \frac{R_n^{(n-1)}(c_{n-1})}{n(n-1)\cdots 2(c_{n-1}-x_0)} \end{aligned}$$

for some  $c_1$  between  $x_0$  and  $x$ ,  $c_2$  between  $x_0$  and  $c_1$ ,  $\dots$ , and  $c_{n-1}$  between  $x_0$  and  $c_{n-2}$ . Then we have

$$\lim_{x \rightarrow x_0} \frac{R_n(x)}{(x-x_0)^n} = \frac{1}{n!} \lim_{c_{n-1} \rightarrow x_0} \frac{R_n^{(n-1)}(c_{n-1}) - R_n^{(n-1)}(x_0)}{c_{n-1} - x_0} = \frac{1}{n!} R_n^{(n)}(x_0) = 0.$$

This proves Theorem 3.4.2.

Now suppose  $f$  has  $(n+1)$ -st order derivative between  $x_0$  and  $x$ . Then we have similar computation

$$\begin{aligned} \frac{R_n(x)}{(x-x_0)^{n+1}} &= \frac{R_n(x) - R_n(x_0)}{(x-x_0)^{n+1} - (x_0-x_0)^{n+1}} = \frac{R'_n(c_1)}{(n+1)(c_1-x_0)^n} \\ &= \cdots = \frac{R_n^{(n)}(c_n)}{(n+1)n(n-1)\cdots 2(c_n-x_0)} \\ &= \frac{R_n^{(n)}(c_n) - R_n^{(n)}(x_0)}{(n+1)!(c_n-x_0)} = \frac{R_n^{(n+1)}(c)}{(n+1)!} = \frac{f^{(n+1)}(c)}{(n+1)!}, \end{aligned}$$

where  $c$  is between  $x_0$  and  $x$ . The last equality is due to that fact that the  $(n+1)$ -st order derivative of the degree  $n$  polynomial  $T_n(x)$  is zero.  $\square$

<sup>17</sup>Joseph-Louis Lagrange, born 1736 in Turin (Italy), died 1813 in Paris (France). In analysis, Lagrange invented calculus of variations, Lagrange multipliers, and Lagrange interpolation. He invented the method of solving differential equations by variation of parameters. In number theory, he proved that every natural number is a sum of four squares. He also transformed Newtonian mechanics into a branch of analysis, now called Lagrangian mechanics.

**Example 3.4.4.** By comparing

$$x^4 = (1 + (x - 1))^4 = 1 + 4(x - 1) + 6(x - 1)^2 + 4(x - 1)^3 + (x - 1)^4$$

with the Taylor expansion formula, we get

$$f'(1) = 4, \quad f''(1) = 6 \cdot 2! = 12, \quad f'''(1) = 4 \cdot 3! = 24, \quad f^{(4)}(1) = 4! = 24,$$

and  $f^{(n)}(1) = 0 \cdot n! = 0$  for  $n > 4$ .

**Example 3.4.5.** Suppose  $f(x)$  has second order derivative at  $x_0$ . Then we have

$$\begin{aligned} f(x_0 + \Delta x) &= f(x_0) + f'(x_0)\Delta x + \frac{f''(x_0)}{2}\Delta x^2 + o(\Delta x^2) \\ f(x_0 + 2\Delta x) &= f(x_0) + 2f'(x_0)\Delta x + 2f''(x_0)\Delta x^2 + o(\Delta x^2). \end{aligned}$$

This implies

$$f(x_0 + 2\Delta x) - 2f(x_0 + \Delta x) + f(x_0) = f''(x_0)\Delta x^2 + o(\Delta x^2),$$

and gives another expression of the second order derivative as a limit

$$f''(x_0) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + 2\Delta x) - 2f(x_0 + \Delta x) + f(x_0)}{\Delta x^2}.$$

Exercises 3.96 and 3.97 extend the example.

**Example 3.4.6.** Suppose  $f(x)$  has second order derivative on  $[0, 1]$ . Suppose  $|f(0)| \leq 1$ ,  $|f(1)| \leq 1$  and  $|f''(x)| \leq 1$ . We would like to estimate the size of  $f'(x)$ .

Fix any  $0 < x < 1$ . By the second order Taylor expansion at  $x$  and the remainder formula, we have

$$\begin{aligned} f(0) &= f(x) + f'(x)(0 - x) + \frac{f''(c_1)}{2}(0 - x)^2, \quad c_1 \in (0, x), \\ f(1) &= f(x) + f'(x)(1 - x) + \frac{f''(c_2)}{2}(1 - x)^2, \quad c_2 \in (x, 1). \end{aligned}$$

Subtracting the two, we get

$$f'(x) = f(1) - f(0) + \frac{f''(c_1)}{2}x^2 - \frac{f''(c_2)}{2}(1 - x)^2.$$

By the assumption on the sizes of  $f(1)$ ,  $f(0)$  and  $f''(x)$ , we then get

$$|f'(x)| \leq 2 + \frac{1}{2}(x^2 + (1 - x)^2) \leq \frac{5}{2}.$$

**Exercise 3.94.** Suppose  $f''(0)$  exists and  $f''(0) \neq 0$ . Prove that in the Mean Value Theorem  $f(x) - f(0) = xf'(\theta x)$ , we have  $\lim_{x \rightarrow 0} \theta = \frac{1}{2}$ . This generalizes the observation in Exercise 3.82.

**Exercise 3.95.** Suppose  $f(x)$  has second order derivative at  $x_0$ . Let  $h$  and  $k$  be small, distinct and nonzero numbers. Find the quadratic function  $q(x) = a + b\Delta x + c\Delta x^2$  satisfying

$$q(x_0) = f(x_0), \quad q(x_0 + h) = f(x_0 + h), \quad q(x_0 + k) = f(x_0 + k).$$

Then prove that  $\lim_{h,k \rightarrow 0} b = f'(x_0)$  and  $\lim_{h,k \rightarrow 0} c = \frac{f''(x_0)}{2}$  as long as  $\frac{h}{h-k}$  is kept bounded. This provides the geometrical interpretation of the quadratic approximation.

**Exercise 3.96.** Find suitable conditions among constants  $a, b, \lambda, \mu$  so that  $\lambda f(x_0 + a\Delta x) + \mu f(x_0 + b\Delta x) + f(x_0) = f''(x_0)\Delta x^2 + o(\Delta x^2)$  holds for functions with second order derivative.

**Exercise 3.97.** For  $h \neq 0$ , we have the *difference operator*  $\Delta = \Delta_h$ , which can be applied to a function  $f(x)$  to produce a new function

$$\Delta f(x) = f(x+h) - f(x).$$

By repeatedly applying the operator, we get the high order difference operator

$$\Delta^n f(x) = \Delta(\Delta^{n-1} f)(x) = \Delta^{n-1}(\Delta f)(x).$$

1. Prove that  $\Delta(af + bg) = a\Delta f + b\Delta g$  and  $\Delta(f(x+b)) = (\Delta f)(x+b)$ .
2. Prove that  $\Delta^n f(x) = \sum_{i=0}^n (-1)^{n-i} \frac{n!}{i!(n-i)!} f(x+ih)$ .
3. Prove that  $\Delta^n(ax+b)^k = 0$  for  $k < n$  and  $\Delta^n(ax+b)^n = n!a^n h^n$ .
4. If  $f^{(n)}(x_0)$  exists, prove that

$$f^{(n)}(x_0) = \lim_{h \rightarrow 0} \frac{\Delta^n f(x_0)}{h^n} = \lim_{h \rightarrow 0} \frac{\Delta^n f(x_0+h)}{h^n}.$$

The last formula for  $f^{(n)}(x_0)$  generalises the formula for  $f''(x_0)$  in Example 3.4.5.

**Exercise 3.98.** Prove that if there is  $M$ , such that  $|f^{(n)}(x)| \leq M$  for all  $n$  and  $x \in [a, b]$ , then the Taylor expansion of  $f(x)$  converges to  $f(x)$  for  $x \in [a, b]$ .

**Exercise 3.99.** Suppose  $f(x)$  has the third order derivative on  $[-1, 1]$ , such that  $f(-1) = 0$ ,  $f(0) = 0$ ,  $f(1) = 1$ ,  $f'(0) = 0$ . Prove that there are  $x \in [-1, 0]$  and  $y \in [0, 1]$ , such that  $f'''(x) + f'''(y) = 6$ .

**Exercise 3.100.** Define high order left and right derivatives. What are some properties of high order one sided derivative?

## Differentiability vs Derivative

Proposition 3.1.5 says that the first order differentiability is equivalent to the existence of the first order derivative. However, the high order differentiability does not necessarily imply the existence of the high order derivative. In other words, the converse of Theorem 3.4.2 is not true.

**Example 3.4.7.** In Example 3.1.3, we studied the first order differentiability (equivalent to first order derivative) of  $|x|^p$ ,  $p > 0$ , at  $x_0 = 0$ . Here we study the high order differentiability and high order derivative.

If  $p$  is an even integer, then  $|x|^p = x^p$  is a polynomial, which is smooth and is differentiable of any order.

If  $p$  is an odd integer, then  $|x|^p = x^p$  for  $x \geq 0$  and  $|x|^p = -x^p$  for  $x \leq 0$ . We will assume  $p = 3$  in the subsequent discussion, and the general case is similar. The function has second order derivative

$$(|x|^3)' = \begin{cases} 3x^2, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -3x^2, & \text{if } x < 0, \end{cases} \quad (|x|^3)'' = \begin{cases} 6x, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -6x, & \text{if } x < 0. \end{cases}$$

By Proposition 3.4.2, the function is second order differentiable at 0, with 0 as the quadratic approximation. Therefore by Example 3.4.2, if  $|x|^3$  is third order differentiable, then the cubic approximation must be of the form  $bx^3$ . However, this implies  $|x|^3 = bx^3 + o(x^3)$ , so that  $\lim_{x \rightarrow 0} \frac{|x|^3}{x^3} = b$  converges. Since this is not true, we conclude that  $|x|^3$  is not third order differentiable at 0. By Proposition 3.4.2 (or by direct verification using the formula for  $(|x|^3)''$ ), the function does not have third order derivative at 0. In general, for odd  $p$ , the function  $|x|^p$  is  $(p-1)$ -st order differentiable at 0 but not  $p$ -th order differentiable. The function also has  $(p-1)$ -st order derivative at 0 but does not have  $p$ -th order derivative.

Next we consider the case  $p$  is not an integer. Again, we take  $p = 2.5$  as an example. The function has second order derivative

$$(|x|^{2.5})' = \begin{cases} 2.5|x|^{1.5}, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -2.5|x|^{1.5}, & \text{if } x < 0, \end{cases} \quad (|x|^{2.5})'' = \begin{cases} 3.75|x|^{0.5}, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -3.75|x|^{0.5}, & \text{if } x < 0. \end{cases}$$

By Proposition 3.4.2, the function is second order differentiable at 0, with 0 as the quadratic approximation. Therefore if  $|x|^{2.5}$  is third order differentiable, then the cubic approximation must be of the form  $bx^3$ . However, this implies  $|x|^{2.5} = bx^3 + o(x^3)$ , so that  $\lim_{x \rightarrow 0} \frac{|x|^{2.5}}{x^3} = b$  converges. Since the limit actually diverges to infinity, we conclude that  $|x|^{2.5}$  is not third order differentiable at 0. By Proposition 3.4.2 (or by direct verification using the formula for  $(|x|^{2.5})''$ ), the function does not have third order derivative at 0. In general, for  $n < p < n+1$ ,  $n$  integer, the function  $|x|^p$  is  $n$ -th order differentiable at 0 but not  $(n+1)$ -st order differentiable. The function also has  $n$ -th order derivative at 0 but does not have  $(n+1)$ -st order derivative.

We note that the  $n$ -order differentiability and the existence of the  $n$ -order derivative are the same for  $|x|^p$ .

**Example 3.4.8.** Suppose  $|f(x)| \leq |x|^p$  for some  $p > n$ ,  $n$  a natural number. Then we have  $f(x) = 0 + 0x + 0x^2 + \cdots + 0x^n + o(x^n)$ . Therefore the function is  $n$ -th order differentiable at 0, with 0 as the approximation.

On the other hand, the function does not even need to be continuous away from 0. An example is given by

$$f(x) = |x|^p D(x) = \begin{cases} |x|^p, & \text{if } x \text{ is rational,} \\ 0, & \text{if } x \text{ is irrational.} \end{cases}$$

Since the only continuity of the function is at 0,  $f'(x)$  does not exist for  $x \neq 0$ . Since the first order derivative is not a function, the second and higher order derivatives are not defined.

**Exercise 3.101.** Determine high order differentiability and the existence of high order derivative at 0.



$$\begin{array}{ll}
1. f(x) = \begin{cases} x^3, & \text{if } x \text{ is rational,} \\ 0, & \text{if } x \text{ is irrational.} \end{cases} & 3. \begin{cases} ax^p, & \text{if } x \geq 0, \\ b(-x)^q, & \text{if } x < 0. \end{cases} \\
2. f(x) = \begin{cases} x^3 \sin \frac{1}{x^2}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases} & 4. \begin{cases} |x|^p \sin \frac{1}{|x|^q}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}
\end{array}$$

### Analytic Function

Suppose a function has high order derivative of any order at  $x_0$ . Then the Taylor expansion  $T_n(x)$  is better and better approximation of  $f(x)$  as  $n$  gets larger and larger. The analogue would be measuring the length of an object by more and more accurate ruler (meter, centimeter, millimeter, micrometer, nanometer, picometer, etc.). The problem is whether such measurement determines the actual length.

The following example gives a function that has all the high order derivatives at 0 to be 0, and yet the function is nonzero. This is analogous to an object that always has length 0 no matter how accurate our ruler is, and yet the object has nonzero length.

**Example 3.4.9 (Cauchy).** Let  $p(x)$  be a polynomial and

$$f(x) = \begin{cases} p\left(\frac{1}{x}\right) e^{-\frac{1}{x^2}}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

At  $x \neq 0$ , we have

$$f'(x) = \left[ -p'\left(\frac{1}{x}\right) \frac{1}{x^2} + p\left(\frac{1}{x}\right) \frac{2}{x^3} \right] e^{-\frac{1}{x^2}} = q\left(\frac{1}{x}\right) e^{-\frac{1}{x^2}},$$

where  $q(x) = -p'(x)x^2 - 2p(x)x^3$  is also a polynomial. Moreover, by  $\lim_{x \rightarrow 0} x^k e^{-\frac{1}{x^2}} = \lim_{x \rightarrow +\infty} x^{-\frac{k}{2}} e^{-x} = 0$  for any  $k$ , we have

$$f'(0) = \lim_{x \rightarrow 0} \frac{f(x)}{x} = \lim_{x \rightarrow 0} \frac{1}{x} p\left(\frac{1}{x}\right) e^{-\frac{1}{x^2}} = 0.$$

Therefore  $f'(x)$  is of the same type as  $f(x)$ , with another polynomial  $q(x)$  in place of  $p(x)$ . In particular, we conclude that

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

has derivatives of all orders, and  $f^{(n)}(0) = 0$  for any  $n$ . This implies that the Taylor expansion of the function is 0 at any order.

**Exercise 3.102.** Prove that the function

$$f(x) = \begin{cases} e^{-\frac{1}{|x|}}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

has all high order derivatives equal to 0.

A function is called *analytic* if it is the limit of its Taylor series. We will see that this is equivalent to that the function is the limit of a power series. The usual functions such as polynomials, power functions, exponential functions, logarithmic functions, trigonometric functions, and inverse trigonometric functions are all analytic. Moreover, the addition, multiplication and composition of analytic functions are still analytic. To truly understand analytic functions (and the claims on what functions are analytic), one needs to study complex analysis.

## 3.5 Application

### Maximum and Minimum

Suppose  $f(x)$  is differentiable at  $x_0$ . By Proposition 3.3.1, a necessary condition for  $x_0$  to be a local extreme is  $f'(x_0) = 0$ . High order approximations can be further used to determine whether  $x_0$  is indeed a local extreme.

**Proposition 3.5.1.** *Suppose  $f(x)$  has  $n$ -th order approximation  $a + b(x - x_0)^n$  at  $x_0$ , with  $b \neq 0$ .*

1. *If  $n$  is odd and  $b \neq 0$ , then  $x_0$  is not a local extreme.*
2. *If  $n$  is even and  $b > 0$ , then  $x_0$  is a local minimum.*
3. *If  $n$  is even and  $b < 0$ , then  $x_0$  is a local maximum.*

*Proof.* We can essentially copy the proof of Proposition 3.3.1. We already know  $a = f(x_0)$ . Fix any  $0 < \epsilon < |b|$ , so that  $b - \epsilon$  and  $b$  have the same sign. Then there is  $\delta > 0$ , such that

$$\begin{aligned} |x - x_0| < \delta &\implies |f(x) - f(x_0) - b(x - x_0)^n| \leq \epsilon |x - x_0|^n \\ &\iff -\epsilon |x - x_0|^n \leq f(x) - f(x_0) - b(x - x_0)^n \leq \epsilon |x - x_0|^n. \end{aligned}$$

Suppose  $n$  is odd and  $b > 0$ , which implies  $b - \epsilon > 0$ . Then for  $x_0 < x < x_0 + \delta$ , we have  $x - x_0 = |x - x_0|$ , and

$$f(x) - f(x_0) \geq b(x - x_0)^n - \epsilon |x - x_0|^n = (b - \epsilon)|x - x_0|^n > 0.$$

Moreover, for  $x_0 - \delta < x < x_0$ , we have  $x - x_0 = -|x - x_0|$ , and (by  $n$  odd)

$$f(x) - f(x_0) \leq b(x - x_0)^n + \epsilon |x - x_0|^n = -(b - \epsilon)|x - x_0|^n < 0,$$

Therefore  $f(x_0)$  is strictly smaller than the right side and strictly bigger than the left side. In particular,  $x_0$  is not a local extreme. The argument for the case  $b < 0$  is similar.

Suppose  $n$  is even and  $b > 0$ , which implies  $b - \epsilon > 0$ . Then for  $|x - x_0| < \delta$ , we have  $(x - x_0)^n = |x - x_0|^n$ , and

$$f(x) - f(x_0) \geq b(x - x_0)^n - \epsilon |x - x_0|^n = (b - \epsilon)|x - x_0|^n \geq 0.$$

Therefore  $x_0$  is a local minimum. The proof for the case  $b < 0$  is similar.  $\square$

Suppose  $f(x)$  has high order derivatives. Then we have high order approximations by the Taylor expansion. If

$$f'(x_0) = f''(x_0) = \cdots = f^{(n-1)}(x_0) = 0, \quad f^{(n)}(x_0) \neq 0,$$

then the proposition says the following.

1. If  $n$  is odd, then  $x_0$  is not a local extreme.
2. If  $n$  is even and  $f^{(n)}(x_0) > 0$ , then  $x_0$  is a local minimum.
3. If  $n$  is even and  $f^{(n)}(x_0) < 0$ , then  $x_0$  is a local maximum.

This is the generalisation of the first and second order derivative tests for local extrema. We emphasise that the real test is Proposition 3.5.1, which does not require the existence of high order derivatives.

**Example 3.5.1.** By

$$\sin x - x = -\frac{1}{6}x^3 + o(x^4),$$

0 is not a local extreme of  $\sin x - x$ . By

$$\begin{aligned} \cos x + \sec x &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + o(x^5) + \frac{1}{1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + o(x^5)} \\ &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + 1 + \left(\frac{1}{2!}x^2 - \frac{1}{4!}x^4\right) + \left(\frac{1}{2!}x^2\right)^2 + o(x^5) \\ &= 2 + \frac{1}{6}x^4 + o(x^5), \end{aligned}$$

0 is a local minimum of  $\cos x + \sec x$ .

**Example 3.5.2.** The function

$$x^2 + x^3 D(x) = \begin{cases} x^2 + x^3, & \text{if } x \text{ is rational,} \\ x^2, & \text{if } x \text{ is irrational,} \end{cases}$$

is first order differentiable at only 0. However, the function second order differentiable at 0, with quadratic approximation  $x^2$ . Therefore 0 is a local minimum.

**Exercise 3.103.** Do the second and third parts of Proposition 3.5.1 hold if  $>$  and  $<$  are replaced by  $\geq$  and  $\leq$ ?

**Exercise 3.104.** Find local extrema.

1.  $6x^{10} - 10x^6$ .
2.  $x \log x$ .
3.  $(x^2 + 1)e^x$ .
4.  $\left(1 + x + \frac{1}{2}x^2\right)e^{-x}$ .
5.  $\left(1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3\right)e^{-x}$ .

**Exercise 3.105.** Computer the Taylor expansion of  $\sin x - \tan x$  up to 3-rd order and determine whether 0 is a local extreme.

**Exercise 3.106.** Let  $m$  be a natural number. Let  $n$  be an odd natural number. Determine the local extrema of  $(x+1)x^{\frac{m}{n}}$ .

**Exercise 3.107.** Determine the local extrema of

$$f(x) = \begin{cases} \frac{1}{x^4}e^{-\frac{1}{x^2}}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

**Exercise 3.108.** Suppose  $g(y)$  has  $n$ -th order approximation  $a+b(y-y_0)^n$  at  $y_0$ , with  $b \neq 0$ . Suppose  $f(x)$  is continuous at  $x_0$  and  $f(x_0) = y_0$ . Find the condition for  $g(f(x))$  to be differentiable at  $x_0$ .

### Convex and Concave

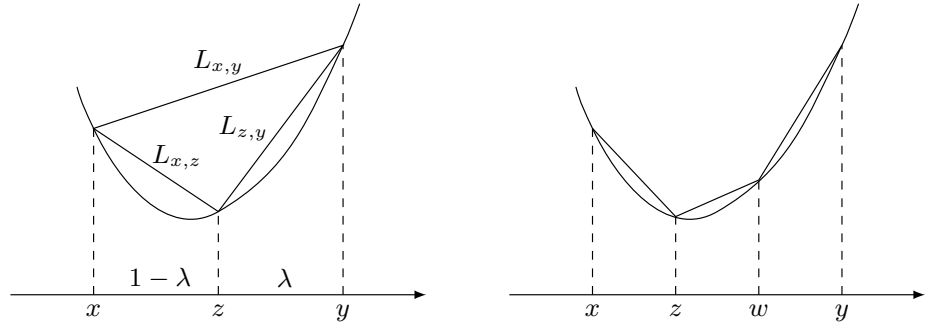
A function is *convex* if the line segment  $L_{x,y}$  connecting any two points  $(x, f(x))$  and  $(y, f(y))$  on the graph of  $f$  lies above the graph of  $f$ . In other words, for any  $x < z < y$ , the point  $(z, f(z))$  lies below  $L_{x,y}$ . This means the inequality

$$L_{x,y}(z) = f(y) + \frac{f(y) - f(x)}{y - x}(z - x) \geq f(z). \quad (3.5.1)$$

The left of Figure 3.5.1 illustrates the definition, and also shows that the convexity is equivalent to any one of the following.

1. slope of  $L_{x,y} \leq$  slope of  $L_{z,y}$ .
2. slope of  $L_{x,z} \leq$  slope of  $L_{x,y}$ .
3. slope of  $L_{x,z} \leq$  slope of  $L_{z,y}$ .

Algebraically, it is not difficult to verify by direct computation that (3.5.1) and the three conditions are equivalent.



**Figure 3.5.1.** *convex function*

The convexity can also be rephrased as follows. Write  $z = \lambda x + (1 - \lambda)y$ . Then  $x < z < y$  is equivalent to  $0 < \lambda < 1$ . Either geometrical consideration or algebraic computation gives

$$L_{x,y}(z) = \lambda f(x) + (1 - \lambda)f(y).$$

Then the convexity means

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \text{ for any } 0 < \lambda < 1. \quad (3.5.2)$$

This can be extended to Jensen<sup>18</sup>'s inequality in Exercise 3.118.

A function is *concave* if the line segment connecting two points on the graph of  $f$  lies below the graph of  $f$ . By exchanging the directions of the inequalities, all the discussions about convex functions can be applied to concave functions.

**Proposition 3.5.2.** *A convex function  $f(x)$  has both one sided derivatives everywhere in the interior of interval, and satisfies*

$$x < y \implies f'_-(x) \leq f'_+(x) \leq f'_-(y) \leq f'_+(y).$$

*Moreover, a convex function is continuous in the interior of interval.*

*Proof.* On the right of Figure 3.5.1, by the third convexity condition, we have

$$\text{slope of } L_{x,z} \leq \text{slope of } L_{z,w} \leq \text{slope of } L_{z,y}.$$

For fixed  $z$ , the inequalities tell us the following.

1. As  $y \rightarrow z^+$ , the slope of  $L_{z,y}$  is decreasing.
2. The slope of  $L_{x,z}$  is a lower bound of  $L_{z,y}$  for all  $y > z$ .

This implies that the right derivative converges and is bounded below by the slope of  $L_{x,z}$

$$f'_+(z) = \lim_{y \rightarrow z^+} (\text{slope of } L_{z,y}) \geq \text{slope of } L_{x,z}.$$

The existence of left derivative can be proved similarly. Then by taking  $x \rightarrow z^-$  on the right in the inequality above, we get

$$f'_+(z) \geq \lim_{x \rightarrow z^-} \text{slope of } L_{x,z} = f'_-(z).$$

To prove the inequality in the proposition, it remains to prove  $x < y$  implying  $f'_+(x) \leq f'_-(y)$ . We note that

$$\text{slope of } L_{x,z} \leq \text{slope of } L_{z,w} \leq \text{slope of } L_{w,y}.$$

By taking  $z \rightarrow x^+$  and  $w \rightarrow y^-$ , we get  $f'_+(x) \leq f'_-(y)$ .

Finally, by Proposition 3.1.3, the existence of one sided derivatives implies the left and right continuity, which is the same as continuity.  $\square$

The converse of Proposition 3.5.2 is also true. The following is the partial converse. The full converse is left as Exercise 3.113.

<sup>18</sup>Johan Jensen, born 1859 in Nakskov (Denmark), died 1925 in Copenhagen (Denmark). He proved the inequality in 1906.

**Proposition 3.5.3.** *Suppose  $f(x)$  is differentiable on an interval. Then  $f(x)$  is convex if and only if  $f'(x)$  is increasing.*

By Propositions 3.3.5, if  $f'(x)$  is also differentiable, then  $f(x)$  is convex if and only if the second order derivative  $f''(x) = (f'(x))' \geq 0$ .

*Proof.* Suppose  $f'$  is increasing and  $x < z < y$ . By the Mean Value Theorem, we have

$$\begin{aligned}\text{slope of } L_{x,z} &= f'(c) \text{ for some } c \in (x, z), \\ \text{slope of } L_{z,y} &= f'(d) \text{ for some } d \in (z, y).\end{aligned}$$

By  $c < d$ , we have  $f'(c) \leq f'(d)$ . This verifies the third condition for convexity.  $\square$

**Example 3.5.3.** The derivative  $(-\log x)' = -\frac{1}{x}$  is decreasing for  $x > 0$ . Therefore  $-\log x$  is a convex function. If  $p, q > 0$  satisfy  $\frac{1}{p} + \frac{1}{q} = 1$ , then by taking  $\lambda, x, y$  to be  $\frac{1}{q}, x^p, y^q$  in (3.5.2), we have

$$\log \left( \frac{1}{p} x^p + \frac{1}{q} y^q \right) \geq \frac{1}{p} \log x^p + \frac{1}{q} \log y^q = \log xy.$$

Taking the exponential, we get the Young's inequality in Exercise 3.74.

**Exercise 3.109.** Are the sum, product, composition, maximum, minimum of two convex functions still convex?

**Exercise 3.110.** Prove that a function  $f(x)$  on an open interval  $(a, b)$  is convex if and only if for any  $a < x < y < b$ , we have  $f(z) \geq L_{x,y}(z)$  for any  $z \in (a, x)$  and  $z \in (y, b)$ .

**Exercise 3.111.** Suppose  $f(x)$  is a convex function. Prove that  $x < y$  implies

$$f'_+(x) \leq \frac{f(y) - f(x)}{y - x} \leq f'_-(y).$$

**Exercise 3.112.** Prove that a function  $f(x)$  on an open interval is convex if and only if for any  $z$ , there is a linear function  $K(x)$  such that  $K(z) = f(z)$  and  $K(x) \leq f(x)$  for all  $x$ .

**Exercise 3.113.** Prove that a function  $f(x)$  on an open interval is convex if and only if  $f(x)$  is left and right differentiable, and

$$x < y \implies f'_-(x) \leq f'_+(x) \leq f'_-(y) \leq f'_+(y).$$

**Exercise 3.114.** Prove that a continuous convex function  $f(x)$  on  $[a, b]$  can be extended to a convex function on  $\mathbb{R}$  if and only if  $f'_+(a)$  and  $f'_-(b)$  are bounded.

**Exercise 3.115.** Prove that a convex function is differentiable at all but countably many places. Exercise 5.138 gives a construction that, for any given countably many places, there is a convex function not differentiable at exactly the given places.

**Exercise 3.116.** Verify the convexity of  $x \log x$  and then use the property to prove the inequality  $(x + y)^{x+y} \leq (2x)^x (2y)^y$ .

**Exercise 3.117.** Suppose  $p \geq 1$ . Show that  $x^p$  is convex for  $x > 0$ . Then for non-negative  $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n$ , take

$$x = \frac{a_i}{(\sum a_i^p)^{\frac{1}{p}}}, \quad y = \frac{b_i}{(\sum b_i^p)^{\frac{1}{p}}}, \quad \lambda = \frac{(\sum a_i^p)^{\frac{1}{p}}}{(\sum a_i^p)^{\frac{1}{p}} + (\sum b_i^p)^{\frac{1}{p}}},$$

in the inequality (3.5.2) and derive Minkowski's inequality in Exercise 3.76.

**Exercise 3.118 (Jensen's Inequality).** Suppose  $f(x)$  is a convex function. For any  $\lambda_1, \lambda_2, \dots, \lambda_n$  satisfying  $\lambda_1 + \lambda_2 + \dots + \lambda_n = 1$  and  $0 < \lambda_i < 1$ , prove that

$$f(\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \dots + \lambda_n f(x_n).$$

Then use this to prove that for  $x_i > 0$ , we have

$$\sqrt[p]{x_1 x_2 \cdots x_n} \leq \frac{x_1 + x_2 + \dots + x_n}{n} \leq \sqrt[p]{\frac{x_1^p + x_2^p + \dots + x_n^p}{n}} \text{ for } p \geq 1,$$

and

$$(x_1 x_2 \cdots x_n)^{\frac{x_1 + x_2 + \dots + x_n}{n}} \leq x_1^{x_1} x_2^{x_2} \cdots x_n^{x_n}.$$

**Exercise 3.119.** Prove that a continuous function on an interval is convex if and only if  $\frac{f(x) + f(y)}{2} \geq f\left(\frac{x+y}{2}\right)$  for any  $x$  and  $y$  on the interval.

## 3.6 Additional Exercise

### Mean Value Theorem for One Sided Derivative

**Exercise 3.120.** Suppose  $f(x)$  is a continuous function on  $[a, b]$  and  $l < \frac{f(b) - f(a)}{b - a}$ .

1. Construct a linear function  $L(x)$  satisfying  $L'(x) = l$  and  $L(a) > f(a)$ ,  $L(b) < f(b)$ .
2. For the linear function  $L(x)$  in the first part, prove that

$$c = \sup\{x \in (a, b) : L(x) \geq f(x) \text{ on } [a, x]\}$$

satisfies  $a < c < b$  and  $L(c) = f(c)$ .

3. Prove that if  $f(x)$  has any one sided derivative at  $c$ , then the one sided derivative is no less than  $l$ .

**Exercise 3.121.** <sup>19</sup> Suppose  $f(x)$  is a continuous function on  $[a, b]$ , such that at any point in  $(a, b)$ , the function is either left or right differentiable. Let  $f'_*(x)$  be one of the one side derivatives at  $x$ . Prove that

$$\inf_{(a,b)} f'_* \leq \frac{f(b) - f(a)}{b - a} \leq \sup_{(a,b)} f'_*.$$

<sup>19</sup>See "Some Remarks on Functions with One-Sided Derivatives" by Miller and Výborný, American Math Monthly **93** (1986) 471-475.

**Exercise 3.122.** Suppose  $f(x)$  is either left or right differentiable at any  $x \in (a, b)$ , and a one sided derivative  $f'_*(x)$  is chosen at every  $x$ . Prove that if  $f'_*(x)$  is increasing, then  $f(x)$  is convex on  $(a, b)$ .

### Existence of High Order Derivative

Exercise 3.89 tells us the condition for a function to be  $(n + k)$ -th order differentiable at  $x_0$  when we already know the function is  $n$ -th order differentiable at  $x_0$ . The high order derivative version of the problem is the following: Suppose  $f^{(n)}(x_0)$  exists, so that we have the Taylor expansion  $T_n(x)$ , and  $f(x) = T_{n-1}(x) + g(x)\Delta x^n$ . (This is the same as  $f(x) = T_n(x) + h(x)\Delta x^n$ ,  $g(x) = \frac{f^{(n)}(x_0)}{n!} + h(x)$ .) What is the condition on  $g(x)$  that corresponds to the existence of  $f^{(n+k)}(x_0)$ ?

**Exercise 3.123.** Suppose  $f''(x_0)$  exists and  $f(x_0) = f'(x_0) = 0$ . Prove that  $f(x) = g(x)(x - x_0)^2$  for function  $g(x)$ , such that  $g(x)$  is continuous at  $x_0$ ,  $g'(x)$  exists for  $x$  near  $x_0$  and  $\neq x_0$ , and  $\lim_{x \rightarrow x_0} g'(x)(x - x_0) = 0$ .

**Exercise 3.124.** Conversely, prove that if  $g(x)$  has the properties in Exercise 3.89, then  $f(x) = g(x)(x - x_0)^2$  has second order derivative at  $x_0$ .

**Exercise 3.125.** Find a continuous function  $g(x)$  that has derivative of any order at any  $x \neq 0$ , yet  $g(x)$  is not differentiable at 0.

**Exercise 3.126.** Extend Exercises 3.123 and 3.124 to high order derivatives.

**Exercise 3.127.** Prove that  $f^{(n+k)}(x_0)$  exists if and only if  $f(x) = T_{n-1}(x) + g(x)\Delta x^n$  for a function  $g(x)$ , such that  $g(x)$  is continuous at  $x_0$ ,  $g^{(k-1)}(x)$  exists for small  $x$  near  $x_0$  and  $x \neq x_0$ , and  $\lim_{x \rightarrow x_0} g^{(i)}(x)(x - x_0)^i = 0$  for  $i = 1, 2, \dots, n - 1$ .

### Relation Between the Bounds of a Function and its Derivatives

In Example 3.4.6, we saw the bounds on a function and its second order derivative impose a bound on the first order derivative. The subsequent exercises provide more examples.

**Exercise 3.128.** Suppose  $f(x)$  is a function on  $[0, 1]$  with second order derivative and satisfying  $f(0) = f'(0) = 0$ ,  $f(1) = 1$ . Prove that if  $f''(x) \leq 2$  for any  $0 < x < 1$ , then  $f(x) = x^2$ . In other words, unless  $f(x) = x^2$ , we will have  $f''(x) > 2$  somewhere on  $(0, 1)$ .

**Exercise 3.129.** Consider functions  $f(x)$  on  $[0, 1]$  with second order derivative and satisfying  $f(0) = f(1) = 0$  and  $\min_{[0,1]} f(x) = -1$ . What would be the “lowest bound” for  $f''(x)$ ? In other words, find the biggest  $a$ , such that any such function  $f(x)$  satisfies  $f''(x) \geq a$  somewhere on  $(0, 1)$ .

**Exercise 3.130.** Study the constraint on the second order derivative for functions on  $[a, b]$  satisfying  $f(a) = A$ ,  $f(b) = B$  and  $\min_{[a,b]} f(x) = m$ .

**Exercise 3.131.** Suppose  $f(x)$  has the second order derivative on  $(a, b)$ . Suppose  $M_0, M_1,$



$M_2$  are the suprema of  $|f(x)|$ ,  $|f'(x)|$ ,  $|f''(x)|$  on the interval. By rewriting the remainder formula as an expression of  $f'$  in terms of  $f$  and  $f''$ , prove that for any  $a < x < b$  and  $0 < h < \max\{x - a, b - x\}$ , we have

$$|f'(x)| \leq \frac{h}{2}M_2 + \frac{2}{h}M_0.$$

Then prove  $M_1 \leq 2\sqrt{M_2M_0}$  in case  $b = +\infty$ . Also verify that the equality happens for

$$f(x) = \begin{cases} 2x^2 - 1, & \text{if } -1 < x < 0, \\ \frac{x^2 - 1}{x^2 + 1}, & \text{if } x \geq 0. \end{cases}$$

**Exercise 3.132.** Suppose  $f(x)$  has the second order derivative on  $(a, +\infty)$ . Prove that if  $f''(x)$  is bounded and  $\lim_{x \rightarrow +\infty} f(x) = 0$ , then  $\lim_{x \rightarrow +\infty} f'(x) = 0$ .

### Cauchy Form of the Remainder

The Lagrange form (3.4.1) is the simplest form of the remainder. However, for certain functions, it is more suitable to use the *Cauchy form* of the remainder

$$R_n(x) = \frac{f^{(n+1)}(c)}{n!}(x - c)^n(x - x_0).$$

The proof makes use of the function defined for any fixed  $x$  and  $x_0$

$$F(t) = f(x) - f(t) - f'(t)(x - t) - \frac{f''(t)}{2}(x - t)^2 - \cdots - \frac{f^{(n)}(t)}{n!}(x - t)^n.$$

**Exercise 3.133.** Prove the Cauchy form by applying the Mean Value Theorem to  $F(t)$  for  $t$  between  $x_0$  and  $x$ .

**Exercise 3.134.** Prove the Lagrange form (3.4.1) by applying Cauchy's Mean Value Theorem to  $F(t)$  and  $G(t) = (x - t)^{n+1}$ .

**Exercise 3.135.** Derive a general formula for the remainder by applying Cauchy's Mean Value Theorem to  $F(t)$  and any  $G(t)$ .

**Exercise 3.136.** Prove that the remainder of the Taylor series of  $(1 + x)^p$  satisfies

$$|R_n| \leq \rho_n = A \left| \frac{p(p-1) \cdots (p-n)}{n!} x^{n+1} \right| \text{ for } |x| < 1,$$

where  $A = (1 + |x|)^{p-1}$  for  $p \geq 1$  and  $A = (1 - |x|)^{p-1}$  for  $p < 1$ . Then use Exercise 1.59 to show that  $\lim_{n \rightarrow \infty} \rho_n = 0$ . This shows that the Taylor series of  $(1 + x)^p$  converges for  $|x| < 1$ .

**Exercise 3.137.** Study the convergence of the Taylor series of  $\log(1 + x)$ .

### Estimation of $\sin x$ and $\cos x$

**Exercise 3.138.** Prove  $x > \sin x > \frac{2}{\pi}x$  for  $0 < x < \frac{\pi}{2}$ .

Exercise 3.139. Let

$$f_k(x) = x - \frac{x^3}{3!} + \cdots + (-1)^{k-1} \frac{x^{2k-1}}{(2k-1)!} - \sin x,$$

$$g_k(x) = 1 - \frac{x^2}{2!} + \cdots + (-1)^k \frac{x^{2k}}{(2k)!} - \cos x.$$

Verify that  $g'_k = -f_k$  and  $f'_{k+1} = g_k$ . Then prove that for  $x > 0$ , we have

$$x - \frac{x^3}{3!} + \cdots - \frac{x^{4k-1}}{(4k-1)!} < \sin x < x - \frac{x^3}{3!} + \cdots - \frac{x^{4k-1}}{(4k-1)!} + \frac{x^{4k+1}}{(4k+1)!}.$$

Also derive the similar inequalities for  $\cos x$ .

Exercise 3.140. For  $0 < x < \frac{\pi}{2}$ , prove that

$$x - \frac{x^3}{3!} + \cdots - \frac{x^{4k-1}}{(4k-1)!} + \frac{2}{\pi} \frac{x^{4k+1}}{(4k+1)!} < \sin x < x - \frac{x^3}{3!} + \cdots + \frac{x^{4k+1}}{(4k+1)!} - \frac{2}{\pi} \frac{x^{4k+3}}{(4k+3)!},$$

and

$$1 - \frac{x^2}{2!} + \cdots - \frac{x^{4k-2}}{(4k-2)!} + \frac{2}{\pi} \frac{x^{4k}}{(4k)!} < \cos x < 1 - \frac{x^2}{2!} + \cdots + \frac{x^{4k}}{(4k)!} - \frac{2}{\pi} \frac{x^{4k+2}}{(4k+2)!}.$$

Exercise 3.141. Let

$$f_n(x) = 1 + x - \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots + s_1(n) \frac{x^n}{n!}, \quad s_1(n) = \begin{cases} 1, & \text{if } n = 4k, 4k+1, \\ -1, & \text{if } n = 4k+2, 4k+3, \end{cases}$$

$$g_n(x) = 1 - x - \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + s_2(n) \frac{x^n}{n!}, \quad s_2(n) = \begin{cases} 1, & \text{if } n = 4k-1, 4k, \\ -1, & \text{if } n = 4k+1, 4k+2. \end{cases}$$

Prove that

$$f_{4k+1}(x) - \sqrt{2} \frac{x^{4k+2}}{(4k+2)!} < \cos x + \sin x < f_{4k-1}(x) + \sqrt{2} \frac{x^{4k}}{(4k)!},$$

and

$$g_{4k}(x) - \sqrt{2} \frac{x^{4k+1}}{(4k+1)!} < \cos x - \sin x < g_{4k+2}(x) + \sqrt{2} \frac{x^{4k+3}}{(4k+3)!}.$$

Moreover, derive similar inequalities for  $a \cos x + b \sin x$ .

### Ratio Rule

By specializing the ratio rule in Exercise 1.58 to  $y_n = l^n$ , we get the limit version of the ratio rule in Exercises 1.59. By making other choices of  $y_n$  and using the linear approximations to estimate the quotient  $\frac{y_{n+1}}{y_n}$ , we get other versions of the ratio rule.

Exercise 3.142. Prove that if  $p > q > r > 0$ , then  $1 - px < (1 - x)^q < 1 - rx$  and  $1 + px > (1 + x)^q > 1 + rx$  for sufficiently small  $x > 0$ . Then prove the following.

1. If  $\left| \frac{x_{n+1}}{x_n} \right| \leq 1 - \frac{p}{n}$  for some  $p > 0$  and big  $n$ , then  $\lim_{n \rightarrow \infty} x_n = 0$ .

2. If  $\left| \frac{x_{n+1}}{x_n} \right| \geq 1 + \frac{p}{n}$  for some  $p > 0$  and big  $n$ , then  $\lim_{n \rightarrow \infty} x_n = \infty$ .

Exercise 3.143. Study the limits.

1.  $\lim_{n \rightarrow \infty} \frac{(n!)^2 a^n}{(2n)!}$ .                      2.  $\lim_{n \rightarrow \infty} \frac{(n+a)^{n+b}}{c^n n!}$ .

Exercise 3.144. Rephrase the rules in Exercise 3.142 in terms of the quotient  $\left| \frac{x_n}{x_{n+1}} \right|$ . Then prove that  $\lim_{n \rightarrow \infty} n \left( \left| \frac{x_n}{x_{n+1}} \right| - 1 \right) > 0$  implies  $\lim_{n \rightarrow \infty} x_n = 0$ . Find the similar condition that implies  $\lim_{n \rightarrow \infty} x_n = \infty$ .

Exercise 3.145. Prove that if  $p > q > r > 0$ , then  $1 - \frac{p}{x \log x} < \frac{(\log(x-1))^q}{(\log x)^q} < 1 - \frac{r}{x \log x}$  and  $1 + \frac{p}{x \log x} > \frac{(\log(x+1))^q}{(\log x)^q} > 1 + \frac{r}{x \log x}$  for sufficiently big  $x > 0$ . Then prove the following.

1. If  $\left| \frac{x_{n+1}}{x_n} \right| \leq 1 - \frac{p}{n \log n}$  for some  $p > 0$  and big  $n$ , then  $\lim_{n \rightarrow \infty} x_n = 0$ .  
 2. If  $\left| \frac{x_{n+1}}{x_n} \right| \geq 1 + \frac{p}{n \log n}$  for some  $p > 0$  and big  $n$ , then  $\lim_{n \rightarrow \infty} x_n = \infty$ .

**Compare  $\left(1 + \frac{1}{x}\right)^{x+p}$  with  $e$**

Exercise 3.146. Prove that if  $p \geq \frac{1}{2}$ , then  $\left(1 + \frac{1}{x}\right)^{x+p}$  is convex and strictly decreasing for  $x > 0$ . In particular, we have  $\left(1 + \frac{1}{x}\right)^{x+p} > e$  for  $p \geq \frac{1}{2}$  and  $x > 0$ .

Exercise 3.147. Prove that if  $p < \frac{1}{2}$ , then  $\left(1 + \frac{1}{x}\right)^{x+p}$  is strictly increasing for  $x > \frac{p}{1-2p}$ . In particular, we have  $\left(1 + \frac{1}{x}\right)^{x+p} < e$  for  $p < \frac{1}{2}$  and  $x > \frac{p}{1-2p}$ .

Exercise 3.148. Use Exercises 3.146 and 3.147 to prove that  $\left(1 + \frac{1}{x}\right)^{x+p} > e$  for all  $x > 0$  if and only if  $p \geq \frac{1}{2}$ .

Exercise 3.149. Find those  $p$  such that  $\left(1 + \frac{1}{x}\right)^{x+p} < e$  for all  $x > 0$ .

1. Convert the problem to  $p < f\left(\frac{1}{x}\right)$  for all  $x > 0$ , with  $f(x) = \frac{1}{\log(1+x)} - \frac{1}{x}$ .  
 2. Use  $\lim_{x \rightarrow +\infty} f(x) = 0$  to show that  $p \leq 0$  is necessary.

3. Use Exercise 3.147 to show that  $p \leq 0$  is also sufficient.

**Exercise 3.150.** Use Exercises 3.146 and 3.147 to prove that for all  $x > 0$ , we have

$$0 < e - \left(1 + \frac{1}{x}\right)^x < e \left(\sqrt{1 + \frac{1}{x}} - 1\right) < \frac{e}{2x}.$$

**Exercise 3.151.** The following compares  $\left(1 + \frac{1}{n}\right)^{n+p}$  with  $e$  for all natural numbers  $n$ .

1. Prove that  $u - \frac{1}{u} > 2 \log u$  for  $u > 1$  and  $u - \frac{1}{u} < 2 \log u$  for  $u < 1$ . Then use the inequality to show that  $f(x)$  in Exercise 3.149 is strictly decreasing for  $x > 0$ .
2. Find the supremum and infimum of  $f(x)$  for  $x > 0$  and recover the conclusions of Exercises 3.146 and 3.149.
3. Prove that  $\left(1 + \frac{1}{n}\right)^{n+p} > e$  for all natural number  $n$  if and only if  $p \geq \frac{1}{2}$ .
4. Prove that  $\left(1 + \frac{1}{n}\right)^{n+p} < e$  for all natural number  $n$  if and only if  $p \leq \frac{1}{\log 2} - 1$ .

## Chapter 4

# Integration

## 4.1 Riemann Integration

Discovered independently by Newton<sup>20</sup> and Leibniz, the integration was originally the method of using the antiderivative to find the area under a curve. So the method started as an application of the differentiation. Then Riemann<sup>21</sup> studied the limiting process leading to the area and established the integration as an independent subject. The new viewpoint further led to other integration theories, among which the most significant is the Lebesgue integration in Section 10.1.

### Riemann Sum

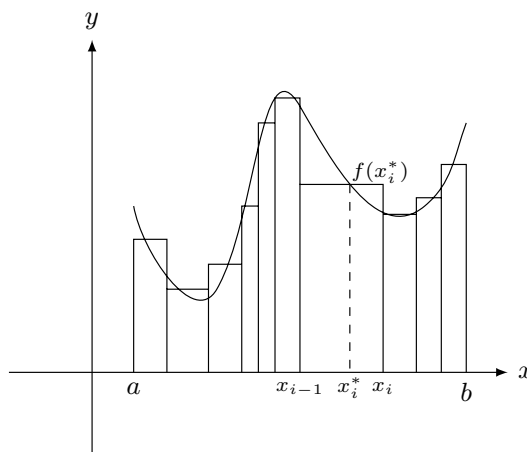
Let  $f(x)$  be a function on a bounded interval  $[a, b]$ . To compute the area of the region between the graph of  $f(x)$  and the  $x$ -axis, we choose a *partition* of the interval

$$P: a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$

Then we approximate the region by a sequence of rectangles with base  $[x_{i-1}, x_i]$  and height  $f(x_i^*)$ , where  $x_i^* \in [x_{i-1}, x_i]$  are the *sample points*. The total area of the rectangles is the *Riemann sum*

$$S(P, f) = \sum_{i=1}^n f(x_i^*)(x_i - x_{i-1}) = \sum_{i=1}^n f(x_i^*)\Delta x_i.$$

Note that  $S(P, f)$  also depends on the choices of  $x_i^*$ , although the choice does not explicitly appear in the notation.



**Figure 4.1.1.** *Riemann sum.*

<sup>20</sup>Isaac Newton, born 1643 in Woolsthorpe (England), died 1727 in London (England). Newton is a giant of science and one of the most influential people in human history. Together with Gottfried Leibniz, he invented calculus. For Newton, differentiation is the fundamental concept, and integration only means to recover a function from its derivative.

<sup>21</sup>Georg Friedrich Bernhard Riemann, born 1826 in Breselenz (Germany), died 1866 in Selasca (Italy).

We expect the Riemann sum to be more accurate approximation of the area as the size of the partition

$$\|P\| = \max_{1 \leq i \leq n} \Delta x_i$$

gets smaller. This leads to the definition of the *Riemann integral*.

**Definition 4.1.1.** A function  $f(x)$  on a bounded interval  $[a, b]$  is *Riemann integrable*, with *integral*  $I$ , if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

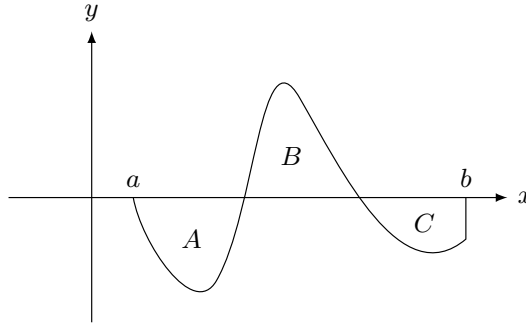
$$\|P\| < \delta \implies |S(P, f) - I| < \epsilon.$$

Due to the similarity to the definition of limits, we write

$$I = \int_a^b f(x) dx = \lim_{\|P\| \rightarrow 0} S(P, f).$$

The numbers  $a$  and  $b$  are called the *lower limit* and the *upper limit* of the integral.

Since the Riemann sum  $S(P, f)$  takes into account of the sign of  $f(x)$ , the integration is actually the *signed* area, which counts the part of the area corresponding to  $f(x) < 0$  as negative.



**Figure 4.1.2.**  $\int_a^b f(x) dx = -\text{area}(A) + \text{area}(B) - \text{area}(C).$

**Example 4.1.1.** For the constant function  $f(x) = c$  on  $[a, b]$  and any partition  $P$ , we have

$$S(P, c) = \sum_{i=1}^n c \Delta x_i = c \sum_{i=1}^n \Delta x_i = c(b - a).$$

Therefore the constant function is Riemann integrable, with  $\int_a^b c dx = c(b - a).$

**Example 4.1.2.** For the constant function  $f(x) = x$  on  $[0, 1]$  and any partition  $P$ , we choose the middle points  $x_i^* = \frac{x_i + x_{i-1}}{2}$  as the sample points. The corresponding Riemann sum

$$S_{\text{mid}}(P, x) = \sum_{i=1}^n x_i^* (x_i - x_{i-1}) = \sum_{i=1}^n \frac{x_i^2 - x_{i-1}^2}{2} = \frac{x_n^2 - x_0^2}{2} = \frac{1}{2}.$$

This suggests  $\int_0^1 x dx = \frac{1}{2}$ . However, in order to rigorously establish the claim, we also need to consider the other choices of sample points. We compare the Riemann sum  $S(P, x)$  of any sample points  $x_i^{**}$  with the Riemann sum  $S_{\text{mid}}(P, x)$  of the middle sample points

$$\left| S(P, x) - \frac{1}{2} \right| = |S(P, x) - S_{\text{mid}}(P, x)| \leq \sum_{i=1}^n |x_i^{**} - x_i^*| \Delta x_i \leq \sum_{i=1}^n \frac{\|P\|}{2} \Delta x_i = \frac{\|P\|}{2}.$$

This rigorously shows that  $x$  is integrable and  $\int_0^1 x dx = \frac{1}{2}$ .

**Example 4.1.3.** Consider the function that is constantly zero except at  $x = c$

$$d_c(x) = \begin{cases} 0, & \text{if } x \neq c, \\ 1, & \text{if } x = c. \end{cases}$$

For any partition  $P$  of a bounded closed interval  $[a, b]$  containing  $c$ , we have

$$S(P, d_c) = \begin{cases} 0, & \text{if } x_i^* = c, \\ \Delta x_k, & \text{if } x_{k-1} < x_k^* = c < x_k, \\ \Delta x_k, & \text{if } x_1^* = a = c, k = 1 \text{ or } x_n^* = b = c, k = n, \\ \Delta x_k + \Delta x_{k+1}, & \text{if } x_k^* = x_{k+1}^* = x_k = c, k \neq 0, k \neq n. \end{cases}$$

This implies  $|S(P, d_c)| \leq 2\|P\|$ . Therefore  $d_c(x)$  is integrable, and  $\int_a^b d_c(x) dx = 0$ .

**Example 4.1.4.** For the Dirichlet function  $D(x)$  in Example 2.1.2 and any partition  $P$  of  $[a, b]$ , we have  $S(P, D) = b - a$  if all  $x_i^*$  are rational numbers and  $S(P, D) = 0$  if all  $x_i^*$  are irrational numbers. Therefore the Dirichlet function is not integrable.

**Example 4.1.5.** Consider Thomae's function  $R(x)$  in Example 2.3.2. For any natural number  $N$ , let  $A_N$  be the set of rational numbers in  $[0, 1]$  with denominators  $\leq N$ . Then  $A_N$  is finite, containing say  $\nu_N$  numbers. For any partition  $P$  of  $[0, 1]$  and choices of  $x_i^*$ , the Riemann sum  $S(P, R)$  can be divided into two parts. The first part consists of those intervals with  $x_i^* \in A_N$ , and the second part has  $x_i^* \notin A_N$ . The number of terms in the first part is  $\leq 2\nu_N$ , where the factor 2 takes into account of the possibility that two intervals sharing the same point in  $A_N$ . Therefore that the total length of the intervals in the first part is  $\leq 2\nu_N\|P\|$ . Moreover, the total length of the intervals in the second part is  $\leq 1$ , the total length of the whole interval  $[0, 1]$ . Since  $0 < R(x_i^*) \leq 1$  in the first part and  $0 \leq R(x_i^*) \leq \frac{1}{N}$  in the second part, we have

$$0 \leq S(P, R) \leq 2\nu_N\|P\| + \frac{1}{N}1 = 2\nu_N\|P\| + \frac{1}{N}.$$

By taking  $\|P\| < \delta = \frac{1}{2N\nu_N}$ , for example, we get  $0 \leq S(P, R) < \frac{2}{N}$ . Since  $N$  can be arbitrarily big, we conclude that the function is integrable on  $[0, 1]$ , with  $\int_0^1 R(x) dx = 0$ .

**Exercise 4.1.** Compute the Riemann sums.



1.  $f(x) = x$ ,  $x_i = \frac{i}{n}$ ,  $x_i^* = \frac{i-1}{n}$ .
2.  $f(x) = x^2$ ,  $x_i = \frac{i}{n}$ ,  $x_i^* = \frac{i}{n}$ .
3.  $f(x) = x^2$ ,  $x_i = \frac{i}{n}$ ,  $x_i^* = \frac{2i-1}{2n}$ .
4.  $f(x) = a^x$ ,  $x_i = \frac{i}{n}$ ,  $x_i^* = \frac{i-1}{n}$ .

**Exercise 4.2.** Determine the integrability. For the integrable ones, find the integrals.

1.  $\begin{cases} 0, & \text{if } 0 \leq x < 1 \\ 1, & \text{if } 1 \leq x \leq 2 \end{cases}$  on  $[0, 2]$ .
2.  $\begin{cases} x, & \text{if } x \text{ is rational} \\ 0, & \text{if } x \text{ is irrational} \end{cases}$  on  $[0, 1]$ .
3.  $\begin{cases} 1, & \text{if } x = \frac{1}{n}, n \in \mathbb{Z} \\ 0, & \text{otherwise} \end{cases}$  on  $[-1, 1]$ .
4.  $\begin{cases} \frac{1}{n}, & \text{if } x = \frac{1}{2^n}, n \in \mathbb{N} \\ 0, & \text{otherwise} \end{cases}$  on  $[0, 1]$ .

**Exercise 4.3.** Compute the Riemann sum of  $x^2$  on  $[0, 1]$  by using the sample points  $x_i^* = \sqrt{\frac{x_i^2 + x_i x_{i-1} + x_{i-1}^2}{3}}$ . Then show that  $\int_0^1 x^2 dx = \frac{1}{3}$ .

**Exercise 4.4.** Suppose a function  $f(x)$  on  $[a, b]$  has the property that for any  $\epsilon > 0$ , there are only finitely many places where  $|f(x)| \geq \epsilon$ . Prove that  $\int_a^b f(x) dx = 0$ .

## Riemann Integrability

Because the Riemann integral is defined as a limit, the Riemann integrability is a convergence problem, and has properties similar to the limit of sequences and functions.

**Proposition 4.1.2.** *Riemann integrable functions are bounded.*

*Proof.* Let  $f(x)$  be integrable on a bounded interval  $[a, b]$  and let  $I$  be the integral. Then for  $\epsilon = 1 > 0$ , there is a partition  $P$ , such that for any choice of  $x_i^*$ , we have

$$\left| \sum_{i=1}^n f(x_i^*) \Delta x_i - I \right| = |S(P, f) - I| < 1.$$

Now we fix  $x_2^*, x_3^*, \dots, x_n^*$ , so that  $\sum_{i=2}^n f(x_i^*) \Delta x_i$  is a fixed bounded number. Then

$$|f(x_1^*) \Delta x_1| \leq \left| \sum_{i=2}^n f(x_i^*) \Delta x_i - I \right| + 1$$

for any  $x_1^* \in [x_0, x_1]$ . This shows that  $\frac{1}{\Delta x_1} (|\sum_{i=2}^n f(x_i^*) \Delta x_i - I| + 1)$  is a bound for  $f(x)$  on the first interval  $[x_0, x_1]$  of the partition. Similar argument shows that the function is bounded on any other interval of the partition. Since the partition contains finitely many intervals, the function is bounded on  $[a, b]$ .  $\square$

The Dirichlet function in Example 4.1.4 shows that the converse of Proposition 4.1.2 is false. A more refined condition for the integrability can be obtained by the Cauchy criterion (see Exercises 2.101 and 2.102 for a very general version of Cauchy criterion). The Riemann sum converges if and only if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|P\|, \|P'\| < \delta \implies |S(P, f) - S(P', f)| < \epsilon. \quad (4.1.1)$$

Note that hidden in the notation is the choices of  $x_i^*$  and  $x_i'^*$  for  $P$  and  $P'$ . For the special case  $P = P'$ , we have

$$S(P, f) - S(P', f) = \sum_{i=1}^n (f(x_i^*) - f(x_i'^*)) \Delta x_i.$$

For fixed  $P = P'$  and all possible choices of  $x_i^*$  and  $x_i'^*$ , the difference above is exactly bounded by

$$\sup_{\text{all } x_i^*} S(P, f) - \inf_{\text{all } x_i^*} S(P, f) = \sum_{i=1}^n \left( \sup_{[x_{i-1}, x_i]} f - \inf_{[x_{i-1}, x_i]} f \right) \Delta x_i.$$

Define the *oscillation* of a bounded function  $f(x)$  on an interval  $[a, b]$  to be

$$\omega_{[a,b]}(f) = \sup_{x,y \in [a,b]} |f(x) - f(y)| = \sup_{[a,b]} f - \inf_{[a,b]} f.$$

Then

$$\sup_{\text{all } x_i^*} S(P, f) - \inf_{\text{all } x_i^*} S(P, f) = \sum_{i=1}^n \omega_{[x_{i-1}, x_i]}(f) \Delta x_i,$$

is the Riemann sum of the oscillations, which we denote by  $\omega(P, f)$ . Then the specialization of the Cauchy criterion (4.1.1) to the case  $P = P'$  becomes

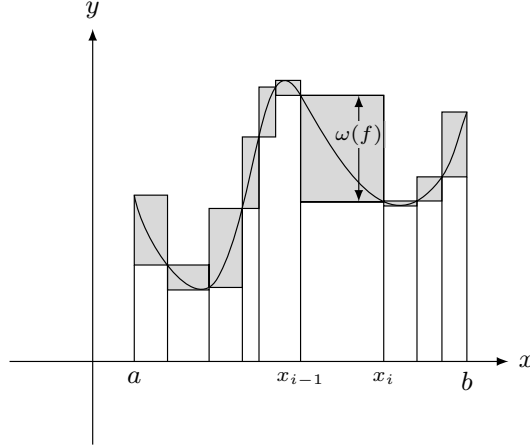
$$\|P\| < \delta \implies \omega(P, f) = \sum_{i=1}^n \omega_{[x_{i-1}, x_i]}(f) \Delta x_i < \epsilon. \quad (4.1.2)$$

The following result says that the specialized Cauchy criterion also implies the general Cauchy criterion (4.1.1).

**Theorem 4.1.3 (Riemann Criterion).** *A bounded function  $f(x)$  on a bounded interval  $[a, b]$  is Riemann integrable if and only if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$  implies  $\omega(P, f) < \epsilon$ .*

*Proof.* As a preparation for the proof, note that for any  $a \leq c \leq b$ , we have

$$\begin{aligned} |f(c)(b-a) - S(P, f)| &= \left| \sum_{i=1}^n (f(c) - f(x_i^*)) \Delta x_i \right| \leq \sum_{i=1}^n |f(c) - f(x_i^*)| \Delta x_i \\ &\leq \sum_{i=1}^n \omega_{[a,b]}(f) \Delta x_i \leq \omega_{[a,b]}(f)(b-a). \end{aligned} \quad (4.1.3)$$



**Figure 4.1.3.** *Riemann sum of oscillations.*

Assume the implication (4.1.2) holds. Let  $P$  and  $P'$  be partitions satisfying  $\|P\|, \|P'\| < \delta$ . Let  $Q = P \cup P'$  be the partition obtained by combining the partition points in  $P$  and  $P'$  together. Take an arbitrary choice of sample points for  $Q$  and form the Riemann sum  $S(Q, f)$ .

The partition  $Q$  is a *refinement* of  $P$  in the sense that it is obtained by adding more partition points to  $P$ . For any interval  $[x_{i-1}, x_i]$  in the partition  $P$ , denote by  $Q_{[x_{i-1}, x_i]}$  the part of  $Q$  lying inside the interval. Then

$$S(Q, f) = \sum_{i=1}^n S(Q_{[x_{i-1}, x_i]}, f).$$

By (4.1.2) and (4.1.3), we have

$$\begin{aligned} |S(P, f) - S(Q, f)| &= \left| \sum_{i=1}^n f(x_i^*) \Delta x_i - \sum_{i=1}^n S(Q_{[x_{i-1}, x_i]}, f) \right| \\ &\leq \sum_{i=1}^n |f(x_i^*)(x_i - x_{i-1}) - S(Q_{[x_{i-1}, x_i]}, f)| \\ &\leq \sum_{i=1}^n \omega_{[x_{i-1}, x_i]}(f) \Delta x_i = \omega(P, f) < \epsilon. \end{aligned}$$

By the same argument, we have  $|S(P', f) - S(Q, f)| < \epsilon$ . Therefore

$$|S(P, f) - S(P', f)| \leq |S(P, f) - S(Q, f)| + |S(P', f) - S(Q, f)| < 2\epsilon. \quad \square$$

The Riemann integrability criterion gives us some important classes of integrable functions.

**Proposition 4.1.4.** *Continuous functions on bounded closed intervals are Riemann integrable.*

*Proof.* By Theorem 2.4.1, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon.$$

If  $\|P\| < \delta$ , then we have

$$x, y \in [x_{i-1}, x_i] \implies |x - y| \leq \|P\| < \delta \implies |f(x) - f(y)| < \epsilon.$$

This implies that the oscillation  $\omega_{[x_{i-1}, x_i]}(f) \leq \epsilon$ , and

$$\omega(P, f) = \sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i \leq \epsilon(b - a).$$

By Theorem 4.1.3, the function is integrable.  $\square$

**Proposition 4.1.5.** *Monotone functions on bounded closed intervals are Riemann integrable.*

*Proof.* Let  $f(x)$  be an increasing function on a bounded closed interval  $[a, b]$ . Then  $\omega_{[x_{i-1}, x_i]}(f) = f(x_i) - f(x_{i-1})$ , and

$$\begin{aligned} \omega(P, f) &= \sum (f(x_i) - f(x_{i-1})) \Delta x_i \\ &\leq \|P\| \sum (f(x_i) - f(x_{i-1})) = \|P\| (f(b) - f(a)). \end{aligned}$$

By Theorem 4.1.3, this implies that the function is integrable.  $\square$

**Example 4.1.6.** The functions  $x$  and  $x^2$  are integrable by Proposition 4.1.4 or 4.1.5.

For the Dirichlet function in Example 2.1.2, we have  $\omega_{[x_{i-1}, x_i]}(D) = 1$ . Therefore  $\omega(P, D) = \sum \Delta x_i = b - a$ , and the function is not integrable by Theorem 4.1.3.

**Example 4.1.7.** Consider the function  $f(x) = \sin \frac{1}{x}$  for  $x \neq 0$  and  $f(0) = 0$ . For any  $\epsilon > 0$ , the function is continuous on  $[\epsilon, 1]$  and is therefore integrable. We claim that the function is actually integrable on  $[0, 1]$ .

Fix any  $\epsilon > 0$ . Applying Proposition 4.1.3 to  $f$  on  $[\epsilon, 1]$ , we find  $\delta > 0$ , such that  $\omega(P', f) < \epsilon$  for partitions  $P'$  of  $[\epsilon, 1]$  satisfying  $\|P'\| < \delta$ . Then for any partition  $P$  of  $[0, 1]$  satisfying  $\|P\| < \min\{\epsilon, \delta\}$ , we have  $x_{j-1} \leq \epsilon < x_j$  for some index  $j$ . Since the partition  $P_\epsilon : \epsilon < x_j < x_{j+1} < \dots < x_n = 1$  of  $[\epsilon, 1]$  satisfies  $\|P_\epsilon\| \leq \|P\| < \delta$ , we have

$$\sum_{i=j+1}^n \omega_{[x_{i-1}, x_i]}(f) \Delta x_i = \omega(P_\epsilon, f) - \omega_{[\epsilon, x_j]}(f)(x_j - \epsilon) \leq \omega(P_\epsilon, f) < \epsilon.$$

On the other hand, by  $x_j \leq x_{j-1} + \|P\| \leq 2\epsilon$  and  $\omega_{[0,1]}f(x) = 2$ , we have

$$\sum_{i=1}^j \omega_{[x_{i-1}, x_i]}(f) \Delta x_i \leq 2 \sum_{i=1}^j \Delta x_i = 2(x_j - 0) \leq 4\epsilon.$$

This implies

$$\omega(P, f) = \sum_{i=1}^j \omega_{[x_{i-1}, x_i]}(f) \Delta x_i + \sum_{i=j+1}^n \omega_{[x_{i-1}, x_i]}(f) \Delta x_i < 5\epsilon.$$

By Proposition 4.1.3, this implies the integrability of  $f$  on  $[0, 1]$ .

Exercise 4.7 gives a general extension of the example.

**Exercise 4.5.** Study the integrability of the functions in Exercise 4.2 again by using Theorem 4.1.3.

**Exercise 4.6.** Suppose  $Q$  is a refinement of  $P$ , prove that  $\omega(Q, f) \leq \omega(P, f)$ .

**Exercise 4.7.** Suppose a bounded function  $f$  on  $[a, b]$  is integrable on  $[a + \epsilon, b]$  for any  $\epsilon > 0$ . Prove that  $f$  is integrable on  $[a, b]$ . In fact, we also have

$$\int_a^b f dx = \lim_{\epsilon \rightarrow 0^+} \int_{a+\epsilon}^b f dx$$

by Theorem 4.4.2.

**Exercise 4.8.** Suppose  $f$  is integrable and  $\inf_{[x_{i-1}, x_i]} f \leq \phi_i \leq \sup_{[x_{i-1}, x_i]} f$ . Prove that

$$\lim_{\|P\| \rightarrow 0} \sum \phi_i \Delta x_i = \int_a^b f dx.$$

**Exercise 4.9.** Suppose  $f(x)$  is a convex function on  $[a, b]$ . By Exercises 3.113,  $f(x)$  is one sided differentiable, and  $f'_-(x), f'_+(x)$  are increasing and therefore integrable. Prove that  $f(b) - f(a) = \int_a^b f'_-(x) dx = \int_a^b f'_+(x) dx$ .

## Integrability of Composition

**Proposition 4.1.6.** Suppose  $f(x)$  is integrable on  $[a, b]$ , and  $\phi(y)$  is bounded and uniformly continuous on the set of values  $f([a, b]) = \{f(x) : x \in [a, b]\}$  of  $f(x)$ . Then the composition  $\phi(f(x))$  is integrable.

The future Theorem 10.4.5 implies that we can drop the requirement that the continuity of  $\phi$  is uniform.

*Proof.* The uniform continuity of  $\phi(y)$  on the values of  $f(x)$  means that, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$y, y' \in f([a, b]), |y - y'| < \delta \implies |\phi(y) - \phi(y')| < \epsilon.$$

By taking  $y = f(x)$  and  $y' = f(x')$ , for any interval  $[c, d] \subset [a, b]$ , we have

$$\begin{aligned} \omega_{[c, d]}(f) < \delta &\implies |f(x) - f(x')| < \delta \text{ for } x, x' \in [c, d] \\ &\implies |\phi(f(x)) - \phi(f(x'))| < \epsilon \text{ for } x, x' \in [c, d] \\ &\implies \omega_{[c, d]}(\phi \circ f) \leq \epsilon. \end{aligned}$$

To prove the integrability of  $\phi(f(x))$ , we decompose the Riemann sum of the oscillations into two parts

$$\omega(P, \phi \circ f) = \sum_{\omega_{[x_{i-1}, x_i]}(f) < \delta} + \sum_{\omega_{[x_{i-1}, x_i]}(f) \geq \delta}.$$

Since  $\omega_{[x_{i-1}, x_i]}(f) < \delta$  implies  $\omega_{[x_{i-1}, x_i]}(\phi \circ f) < \epsilon$ , we may estimate the first part

$$\sum_{\omega_{[x_{i-1}, x_i]}(f) < \delta} \omega_{[x_{i-1}, x_i]}(\phi \circ f) \Delta x_i \leq \sum_{\omega_{[x_{i-1}, x_i]}(f) < \delta} \epsilon \Delta x_i \leq \sum \epsilon \Delta x_i = (b-a)\epsilon.$$

To estimate the second part, we use the integrability of  $f(x)$ . By Theorem 4.1.3, for any  $\epsilon' > 0$ , there is  $\delta' > 0$ , such that

$$\|P\| < \delta' \implies \omega(P, f) < \epsilon'.$$

Then the total length of those intervals with  $\omega_{[x_{i-1}, x_i]}(f) \geq \delta$  can be estimated

$$\delta \sum_{\omega_{[x_{i-1}, x_i]}(f) \geq \delta} \Delta x_i \leq \sum_{\omega_{[x_{i-1}, x_i]}(f) \geq \delta} \omega_{[x_{i-1}, x_i]}(f) \Delta x_i \leq \omega(P, f) \leq \epsilon'.$$

If  $\phi$  is bounded by  $B$ , then  $\omega_{[x_{i-1}, x_i]}(\phi \circ f) \leq 2B$ , and we have

$$\sum_{\omega_{[x_{i-1}, x_i]}(f) \geq \delta} \omega_{[x_{i-1}, x_i]}(\phi \circ f) \Delta x_i \leq 2B \sum_{\omega_{[x_{i-1}, x_i]}(f) \geq \delta} \Delta x_i \leq 2B \frac{\epsilon'}{\delta}.$$

Therefore if we choose  $\epsilon' = \delta\epsilon$  in the first place, then we get

$$\|P\| < \delta' \implies \sum_{\omega_{[x_{i-1}, x_i]}(f) \geq \delta} \omega_{[x_{i-1}, x_i]}(\phi \circ f) \Delta x_i \leq 2B\epsilon.$$

Combining the two estimations together, we get

$$\|P\| < \delta' \implies \omega(P, \phi \circ f) \leq (b-a)\epsilon + 2B\epsilon.$$

Since  $a, b, B$  are all fixed constants, by Theorem 4.1.3, this implies that  $\phi \circ f$  is integrable.  $\square$

**Example 4.1.8.** Suppose  $f(x)$  is integrable and  $\phi(y)$  is continuous on the whole  $\mathbb{R}$ . By Proposition 4.1.2, the values of  $f(x)$  lie in a bounded closed interval. Then by Theorems 2.4.1 and 2.4.2,  $\phi(y)$  is bounded and uniformly continuous on the bounded closed interval. Therefore we may apply Theorem 4.1.6 to conclude that  $\phi(f(x))$  is integrable.

For example, if  $f(x)$  is integrable, then  $f(x)^2$  and  $|f(x)|$  are integrable. Moreover, if  $f(x) \geq 0$  is integrable, then  $\sqrt{f(x)}$  is integrable.

If  $f(x)$  is integrable, and  $|f(x)| > c > 0$  for a constant  $c$ , then the values of  $f$  lie in  $[-B, -c] \cup [c, B]$ , where  $B$  is the bound for  $f$ . Since  $\frac{1}{y}$  is bounded and uniformly continuous on  $[-B, -c] \cup [c, B]$ , we see that  $\frac{1}{f(x)}$  is integrable.

**Example 4.1.9.** In Examples 4.1.3 and 4.1.5, we showed that the function  $d_0$  and Thomae's function  $R$  in Example 2.3.2 are integrable. However, by reason similar to Examples 4.1.4 and 4.1.6, the composition

$$d_0(R(x)) = \begin{cases} 0, & \text{if } x \text{ is rational} \\ 1, & \text{if } x \text{ is irrational} \end{cases} = 1 - D(x)$$

is not integrable.

**Exercise 4.10.** Does the integrability of  $|f(x)|$  imply the integrability of  $f(x)$ ? What about  $f(x)^2$ ? What about  $f(x)^3$ ?

**Exercise 4.11.** Suppose  $\phi(x)$  satisfies  $A(x' - x) \leq \phi(x') - \phi(x) \leq B(x' - x)$  for some constants  $A, B > 0$  and all  $a \leq x < x' \leq b$ .

1. Prove that  $\omega_{[x, x']}(f \circ \phi) = \omega_{[\phi(x), \phi(x')]}(f)$ .
2. Prove that if  $f(y)$  is integrable on  $[\phi(a), \phi(b)]$ , then  $f(\phi(x))$  is integrable on  $[a, b]$ .

Moreover, prove that if  $\phi(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ , satisfying  $A < \phi'(x) < B$  for all  $x \in (a, b)$ , then  $A(x' - x) \leq \phi(x') - \phi(x) \leq B(x' - x)$  for some constants  $A, B > 0$  and all  $a \leq x < x' \leq b$ .

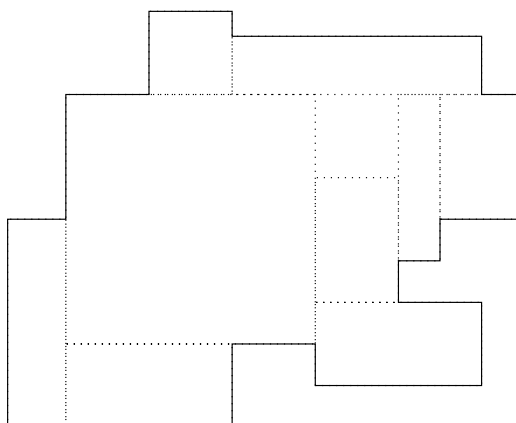
## 4.2 Darboux Integration

Instead of Riemann sum, the integration can also be introduced by directly considering the concept of area. We establish the theory of area by writing down the obvious properties that any reasonable definition of area must satisfy.

1. Bigger subsets have bigger area:  $X \subset Y$  implies  $\mu(X) \leq \mu(Y)$ .
2. Areas can be added: If  $\mu(X \cap Y) = 0$ , then  $\mu(X \cup Y) = \mu(X) + \mu(Y)$ .
3. Rectangles have the usual area:  $\mu(\langle a, b \rangle \times \langle c, d \rangle) = (b - a)(d - c)$ .

Here  $\mu$  (Greek alphabet for  $m$ , used here for *measure*) denotes the area, and we only consider bounded subsets  $X$  of  $\mathbb{R}^2$ . Moreover,  $\langle a, b \rangle$  can be any one of  $(a, b)$ ,  $[a, b]$ ,  $(a, b]$ , or  $[a, b)$ . Again we only consider area of bounded subsets  $X$ .

If a region  $A \subset \mathbb{R}^2$  is a union of finitely many rectangles, then we have  $A = \cup_{i=1}^n I_i$ , such that the intersections between  $I_i$  are at most lines. Since lines have zero area by the third property, we may use the second property to easily calculate the area  $\mu(A) = \sum_{i=1}^n \mu(I_i)$ . We give such a plane region the temporary name “good region”.



**Figure 4.2.1.** *Good region.*

Next we try to approximate a bounded subset  $X$  by good regions, from inside as well as from outside. In other words, we consider good regions  $A$  and  $B$  satisfying  $A \subset X \subset B$ . Then by the first property, the areas must satisfy

$$\mu(A) \subset \mu(X) \subset \mu(B).$$

We already know how to calculate  $\mu(A)$  and  $\mu(B)$ , and yet to calculate  $\mu(X)$ . So we introduce the *inner area*

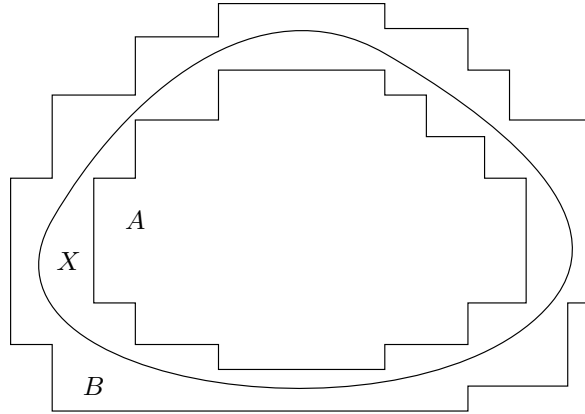
$$\mu_*(X) = \sup\{\mu(A) : A \subset X, A \text{ is a good region}\},$$

as the lower bound for  $\mu(X)$ , and the *outer area*

$$\mu^*(X) = \inf\{\mu(B) : B \supset X, B \text{ is a good region}\},$$

as the upper bound for  $\mu(X)$ .

**Definition 4.2.1.** A bounded subset  $X \subset \mathbb{R}^2$  has area (or *Jordan measurable*) if  $\mu_*(X) = \mu^*(X)$ , and the common value is the *area*  $\mu(X)$  of  $X$ . If  $\mu_*(X) \neq \mu^*(X)$ , then we say  $X$  has no area.



**Figure 4.2.2.** Approximation by good regions.

By the second and third properties in Proposition 1.4.3, we always have  $\mu_*(X) \leq \mu^*(X)$ , and the equality holds if and only if for any  $\epsilon > 0$ , there are good regions  $A$  and  $B$ , such that  $A \subset X \subset B$  and  $\mu(B) - \mu(A) < \epsilon$ . In other words, we can find good inner and outer approximations, such that the difference between the approximations can be arbitrarily small.

**Example 4.2.1.** Consider the triangle with vertices  $(0,0)$ ,  $(1,0)$  and  $(1,1)$ . We partition the interval  $[0,1]$  into  $n$  parts of equal length  $\frac{1}{n}$  and get the inner and outer approximations

$$A_n = \cup_{i=1}^n \left[ \frac{i-1}{n}, \frac{i}{n} \right] \times \left[ 0, \frac{i-1}{n} \right], \quad B_n = \cup_{i=1}^n \left[ \frac{i-1}{n}, \frac{i}{n} \right] \times \left[ 0, \frac{i}{n} \right].$$

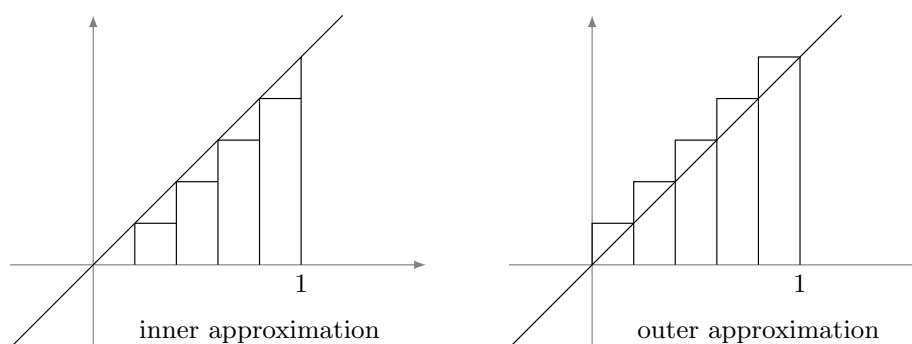


They have area

$$\mu(A_n) = \sum_{i=1}^n \frac{1}{n} \frac{i-1}{n} = \frac{1}{2n}(n-1), \quad \mu(B_n) = \sum_{i=1}^n \frac{1}{n} \frac{i}{n} = \frac{1}{2n}(n+1).$$

By taking sufficiently big  $n$ , the difference  $\mu(B_n) - \mu(A_n) = \frac{1}{n}$  can be arbitrarily small.

Therefore the triangle has area, and the area is  $\lim_{n \rightarrow \infty} \mu(A_n) = \lim_{n \rightarrow \infty} \mu(B_n) = \frac{1}{2}$ .



**Figure 4.2.3.** *Approximating triangle.*

**Example 4.2.2.** For an example of subsets without area, i.e., satisfying  $\mu_*(X) \neq \mu^*(X)$ , let us consider the subset  $X = (\mathbb{Q} \cap [0, 1])^2$  of all rational pairs in the unit square.

Since the only rectangles contained in  $X$  are single points, we have  $\mu(A) = 0$  for any good region  $A \subset X$ . Therefore  $\mu_*(X) = 0$ .

On the other hand, if  $B$  is a good region containing  $X$ , then  $B$  must almost contain the whole square  $[0, 1]^2$ , with the only exception of finitely many horizontal or vertical line segments. Therefore we have  $\mu(B) \geq \mu([0, 1]^2) = 1$ . This implies  $\mu^*(X) \geq 1$  (show that  $\mu^*(X) = 1$ !).

**Exercise 4.12.** Explain that  $\mu^*(X) = 0$  implies  $X$  has zero area.

**Exercise 4.13.** Explain that finitely many points and straight line segments have zero area.

**Exercise 4.14.** Prove that  $X \subset Y$  implies  $\mu_*(X) \leq \mu_*(Y)$  and  $\mu^*(X) \leq \mu^*(Y)$ . In particular, we have  $\mu(X) \leq \mu(Y)$  in case both  $X$  and  $Y$  have areas.

## Darboux Sum

A function  $f$  is *Darboux integrable* if region between the graph of  $f$  and the  $x$ -axis

$$G_{[a,b]}(f) = \{(x, y) : a \leq x \leq b, y \text{ is between } 0 \text{ and } f(x)\}$$

has area. Let

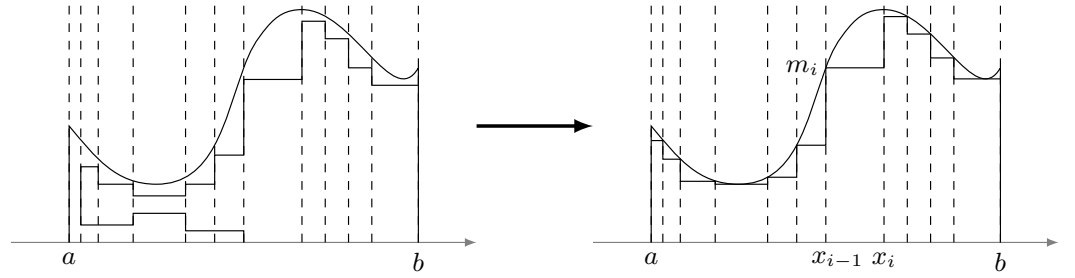
$$H_+ = \{(x, y) : y \geq 0\}, \quad H_- = \{(x, y) : y \leq 0\}$$

be the upper and lower half planes. Then we expect the intersections  $G_{[a,b]}(f) \cap H_+$  and  $G_{[a,b]}(f) \cap H_-$  also have areas, and then define the Darboux integral to be

$$\int_a^b f(x)dx = \mu(G_{[a,b]}(f) \cap H_+) - \mu(G_{[a,b]}(f) \cap H_-).$$

We use the same notation for the Darboux and Riemann integrals, because we will show that the two integrals are equivalent.

For  $f \geq 0$ , Figure 4.2.4 shows that, for any inner approximation, we can always choose “full vertical strips” to get better (meaning larger) inner approximations for  $G_{[a,b]}(f)$ . The same argument applies to outer approximations (here better means smaller). Therefore we only need to consider the approximations by full vertical strips.



**Figure 4.2.4.** Better inner approximations by vertical strips.

An approximation by full vertical strips is determined by a partition  $P$  of the interval  $[a, b]$ . On the  $i$ -th interval  $[x_{i-1}, x_i]$ , the inner strip has height  $m_i = \inf_{[x_{i-1}, x_i]} f$ , and the outer strip has height  $M_i = \sup_{[x_{i-1}, x_i]} f$ . Therefore the inner and outer approximations are

$$A_P = \cup_{i=1}^n [x_{i-1}, x_i] \times [0, m_i] \subset X \subset B_P = \cup_{i=1}^n [x_{i-1}, x_i] \times [0, M_i].$$

The areas of  $A_P$  and  $B_P$  are the *lower and upper Darboux sums*

$$\begin{aligned} L(P, f) &= \mu(A_P) = \sum_{i=1}^n m_i \Delta x_i = \sum_{i=1}^n \left( \inf_{[x_{i-1}, x_i]} f \right) \Delta x_i = \inf_{x_i^*} S(P, f), \\ U(P, f) &= \mu(B_P) = \sum_{i=1}^n M_i \Delta x_i = \sum_{i=1}^n \left( \sup_{[x_{i-1}, x_i]} f \right) \Delta x_i = \sup_{x_i^*} S(P, f). \end{aligned}$$

Then the inner and outer areas of  $G_{[a,b]}(f)$  are the *lower and upper Darboux integrals*

$$\int_a^b f(x)dx = \sup_{\text{all } P} L(P, f), \quad \overline{\int_a^b f(x)dx} = \inf_{\text{all } P} U(P, f),$$

and the Darboux integrability means that the two Darboux integrals are equal.

The concept of Darboux sum and Darboux integral can also be defined for any bounded (not necessarily non-negative) function. We also note that

$$U(P, f) - L(P, f) = \sum_{i=1}^n (M_i - m_i)(x_i - x_{i-1}) = \sum_{i=1}^n \omega_{[x_{i-1}, x_i]}(f) \Delta x_i = \omega(P, f).$$

**Proposition 4.2.2.** *If  $Q$  is a refinement of  $P$ , then*

$$L(P, f) \leq L(Q, f) \leq U(Q, f) \leq U(P, f).$$

*Proof.* We have

$$U(P, f) = \sum \left( \sup_{[x_{i-1}, x_i]} f \right) \Delta x_i \leq \sum \left( \sup_{[a, b]} f \right) \Delta x_i = \left( \sup_{[a, b]} f \right) (b - a),$$

For any interval  $[x_{i-1}, x_i]$  in the partition  $P$ , we apply the inequality above to the part  $Q_{[x_{i-1}, x_i]}$  of  $Q$  lying inside the interval (the notation is from the proof of Theorem 4.1.3), and get

$$U(Q_{[x_{i-1}, x_i]}, f) \leq \left( \sup_{[x_{i-1}, x_i]} f \right) \Delta x_i.$$

This implies

$$U(Q, f) = \sum U(Q_{[x_{i-1}, x_i]}, f) \leq \sum \left( \sup_{[x_{i-1}, x_i]} f \right) \Delta x_i = U(P, f).$$

The inequality  $L(Q, f) \geq L(P, f)$  can be proved similarly. □

**Theorem 4.2.3.** *Suppose  $f$  is a bounded function on a bounded interval  $[a, b]$ . The following are equivalent.*

1. *The function  $f$  is Darboux integrable.*
2. *There is  $I$  with the following property: For any  $\epsilon > 0$ , there is a partition  $P$ , such that  $Q$  refines  $P$  implies  $|S(Q, f) - I| < \epsilon$ .*
3. *For any  $\epsilon > 0$ , there is a partition  $P$ , such that  $\omega(P, f) < \epsilon$ .*

*Proof.* Suppose  $f$  is Darboux integrable. Let

$$\sup_{\text{all } P} L(P, f) = \int_a^b f(x) dx = I = \overline{\int_a^b f(x) dx} = \inf_{\text{all } P} U(P, f).$$

Then for any  $\epsilon > 0$ , there are partitions  $P_1, P_2$ , such that

$$I - \epsilon < L(P_1, f) \leq I, \quad I + \epsilon > U(P_2, f) \geq I.$$

If  $Q$  refines  $P = P_1 \cup P_2$ , then  $Q$  refines  $P_1$  and  $P_2$ . By Proposition 4.2.2, we get

$$I - \epsilon < L(P_1, f) \leq L(Q, f) \leq S(Q, f) \leq U(Q, f) \leq U(P_2, f) < I + \epsilon.$$

This means  $|S(Q, f) - I| < \epsilon$  and proves the second statement.

Suppose we have the second statement. For any  $\epsilon > 0$ , we apply the statement to  $Q = P$ . Then for any choice of partition points  $x_i^*$ , we have

$$I - \epsilon < S(P, f) < I + \epsilon.$$

This implies

$$\omega(P, f) = U(P, f) - L(P, f) = \sup_{x_i^*} S(P, f) - \inf_{x_i^*} S(P, f) \leq (I + \epsilon) - (I - \epsilon) = 2\epsilon.$$

This proves the third statement.

Suppose we have the third statement. Then for any partitions  $Q$  that refines  $P$ , by Proposition 4.2.2, we have

$$U(Q, f) - L(Q, f) \leq U(P, f) - L(P, f) = \omega(P, f) < \epsilon.$$

This implies

$$0 \leq \overline{\int_a^b f(x)dx} - \underline{\int_a^b f(x)dx} = \inf_Q U(Q, f) - \sup_Q L(Q, f) < \epsilon.$$

Since  $\epsilon$  is arbitrary, we conclude that the upper and lower Darboux integrals are equal. This proves the first statement.  $\square$

We note that  $U(P, f) - L(P, f)$  is the area of the good region  $\cup_{i=1}^n [x_{i-1}, x_i] \times [m_i, M_i]$ . The good region can be regarded as an outer approximation of the graph of  $f$ . Therefore Theorem 4.2.3 basically means that the Darboux integrability is equivalent to that the graph of the function has zero area. A vast generalisation of this observation is the third property in Proposition 11.5.2.

**Exercise 4.15.** Use Proposition 4.2.2 to derive Exercise 4.6.

**Exercise 4.16.** Prove that a function is Riemann integrable if and only if  $\lim_{\|P\| \rightarrow 0} U(P, f) = \lim_{\|P\| \rightarrow 0} L(P, f)$ . Moreover, the value of the limit is the Riemann integral.

## Riemann v.s. Darboux

Both Riemann and Darboux integrals are the limit of Riemann sum

$$\int_a^b f(x)dx = I = \lim_P S(P, f),$$

in the sense that, for any  $\epsilon > 0$ , there is  $P$ , such that

$$Q \geq P \implies |S(Q, f) - I| < \epsilon.$$

The detail is in meaning of order  $\geq$  among the partitions. In the definition of Riemann integral, the order is defined in terms of the size  $\|P\|$

$$Q \geq_{\text{Riemann}} P \iff \|Q\| \leq \|P\|.$$

By (second part of) Theorem 4.2.3, the order for Darboux integral is in terms of refinement

$$Q \geq_{\text{Darboux}} P \iff Q \text{ refines } P.$$

The convergence of the limit is the integrability. The criterion for Riemann integrability is given by Theorem 4.1.3, which is  $\omega(P, f) < \epsilon$  for *all*  $P$  satisfying  $\|P\| < \delta$ . The criterion for Darboux integrability is given by (third part of) Theorem 4.2.3, which is  $\omega(P, f) < \epsilon$  for *one*  $P$ . Therefore the Riemann integrability implies the Darboux integrability. It turns out the converse is also true.

**Theorem 4.2.4.** *A bounded function on a bounded interval is Riemann integrable if and only if it is Darboux integrable.*

*Proof.* We need to prove that the Darboux integrability implies the Riemann integrability. Specifically, for any  $\epsilon > 0$ , let a partition  $P$  satisfy  $\omega(P, f) < \epsilon$ . For any partition  $Q$ , the common refinement  $P \cup Q$  is obtained by adding points of  $P$  to  $Q$ . If  $P$  has  $p$  partition points, then  $P \cup Q$  differs from  $Q$  on at most  $p$  intervals of  $Q$ , and the total length of such intervals in  $Q$  and in  $P \cup Q$  together is no more than  $2p\|Q\|$ . On the other hand, if  $f$  is bounded by  $B$  on the whole interval, then the oscillation of  $f$  on any smaller interval is  $\leq 2B$ . Therefore

$$|\omega(Q, f) - \omega(P \cup Q, f)| \leq 2B2p\|Q\|.$$

On the other hand, the refinement  $P \cup Q$  of  $P$  implies (see Exercise 4.6)

$$\omega(P \cup Q, f) \leq \omega(P, f) < \epsilon.$$

Therefore we conclude

$$\omega(Q, f) \leq \omega(P \cup Q, f) + 4Bp\|Q\| < \epsilon + 4Bp\|Q\|.$$

This implies (note that  $B$  and  $p$  are fixed)

$$\|Q\| < \delta = \frac{\epsilon}{4Bp} \implies \omega(Q, f) < 2\epsilon.$$

This proves the Riemann criterion. □

**Exercise 4.17.** Prove that

$$\overline{\int_a^b f(x)dx} = \lim_P U(P, f), \quad \underline{\int_a^b f(x)dx} = \lim_P L(P, f),$$

in either sense of the order among partitions.

The theory of area can be easily extended to the theory of volume for subsets in  $\mathbb{R}^n$ . The details are given in Section 11.4. A byproduct is the multivariable Riemann integral on subsets of Euclidean spaces. In contrast, the definition of multivariable Riemann integral in terms of Riemann sum is more cumbersome.

The underlying idea of Darboux's approach is area. This will lead to the modern measure theory, to be developed in Chapters 9, 10, 11, 12. In this sense, Darboux's approach is superior to Riemann.

On the other hand, Darboux's approach relies on the supremum and infimum, and therefore cannot be applied to the cases where there is no order (like vector valued functions) or the order is questionable (like Riemann-Stieltjes integral).

### 4.3 Property of Riemann Integration

Defined as a certain type of limit, the integration has properties analogous to the limit of sequences and functions.

**Proposition 4.3.1.** *Suppose  $f(x)$  and  $g(x)$  are integrable on  $[a, b]$ . Then  $f(x) + g(x)$  and  $cf(x)$  are also integrable on  $[a, b]$ , and*

$$\begin{aligned}\int_a^b (f(x) + g(x))dx &= \int_a^b f(x)dx + \int_a^b g(x)dx, \\ \int_a^b cf(x)dx &= c \int_a^b f(x)dx.\end{aligned}$$

*Proof.* Let  $I = \int_a^b f(x)dx$  and  $J = \int_a^b g(x)dx$ . For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|P\| < \delta \implies |S(P, f) - I| < \epsilon, \quad |S(P, g) - J| < \epsilon.$$

On the other hand, for the same partition  $P$  and the same sample points  $x_i^*$  for both  $f(x)$  and  $g(x)$ , we have

$$\begin{aligned}S(P, f + g) &= \sum (f(x_i^*) + g(x_i^*))\Delta x_i \\ &= \sum f(x_i^*)\Delta x_i + \sum g(x_i^*)\Delta x_i = S(P, f) + S(P, g).\end{aligned}$$

Therefore

$$\|P\| < \delta \implies |S(P, f + g) - I - J| \leq |S(P, f) - I| + |S(P, g) - J| < 2\epsilon.$$

This shows that  $f + g$  is also integrable, and  $\int_a^b (f(x) + g(x))dx = I + J$ . The proof of  $\int_a^b cf(x)dx = cI$  is similar. □

**Example 4.3.1.** Suppose  $f(x)$  and  $g(x)$  are integrable. By Proposition 4.3.1,  $f(x) + g(x)$  is integrable. By Proposition 4.1.6,  $f(x)^2$ ,  $g(x)^2$  and  $(f(x) + g(x))^2$  are also integrable. Then by Proposition 4.3.1 again, the product

$$f(x)g(x) = \frac{1}{2} [(f(x) + g(x))^2 - f(x)^2 - g(x)^2]$$

is also integrable. However, there is no formula expressing the integral of  $f(x)g(x)$  in terms of the integrals of  $f(x)$  and  $g(x)$ .

Moreover, if  $|g(x)| > c > 0$  for a constant  $c$ , then by the discussion in Example 4.1.8, the function  $\frac{1}{g(x)}$  is integrable, so that the quotient  $\frac{f(x)}{g(x)} = f(x) \frac{1}{g(x)}$  is also integrable.

A vast generalization of the example is given by Proposition 11.5.9, which basically says that a uniformly continuous combination of Riemann integrable functions is Riemann integrable. Specifically, the product function  $\phi(y, z) = yz$  is uniformly continuous on any bounded region of the plane. By the proposition, if  $f(x)$  and  $g(x)$  are Riemann integrable, then the composition  $\phi(f(x), g(x)) = f(x)g(x)$  is Riemann integrable.

**Example 4.3.2.** Suppose  $f(x)$  is integrable on  $[a, b]$ . Suppose  $g(x) = f(x)$  for all  $x$  except at  $c \in [a, b]$ . Then  $g(x) = f(x) + \lambda d_c(x)$ , where  $d_c(x)$  is the function in Example 4.1.3 and  $\lambda = g(c) - f(c)$ . By Example 4.1.3,  $g(x)$  is also integrable and

$$\int_a^b g(x)dx = \int_a^b f(x)dx + \lambda \int_a^b d_c(x)dx = \int_a^b f(x)dx.$$

The example shows that changing an integrable function at finitely many places does not change the integrability and the integral. In particular, it makes sense to talk about the integrability of a function  $f(x)$  on a bounded open interval  $(a, b)$  because any numbers may be assigned as  $f(a)$  and  $f(b)$  without affecting the integrability of (the extended)  $f(x)$  on  $[a, b]$ .

**Exercise 4.18.** Prove that if  $f(x)$  and  $g(x)$  are integrable, then  $\max\{f(x), g(x)\}$  is also integrable.

**Exercise 4.19.** Suppose  $f(x)$  is integrable on  $[a, b]$ . Suppose  $c_n \in [a, b]$  converges. Prove that if  $g(x)$  is obtained by modifying the values of  $f(x)$  at  $c_n$ , and  $g(x)$  is still bounded, then  $g(x)$  is Riemann integrable and  $\int_a^b g(x)dx = \int_a^b f(x)dx$ .

**Proposition 4.3.2.** Suppose  $f(x)$  and  $g(x)$  are integrable on  $[a, b]$ . If  $f(x) \leq g(x)$ , then

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx.$$

*Proof.* Let  $I = \int_a^b f(x)dx$  and  $J = \int_a^b g(x)dx$ . For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|P\| < \delta \implies |S(P, f) - I| < \epsilon, |S(P, g) - J| < \epsilon.$$

By choosing the same  $x_i^*$  for both functions, we have

$$I - \epsilon < S(P, f) = \sum f(x_i^*)\Delta x_i \leq \sum g(x_i^*)\Delta x_i = S(P, g) < J + \epsilon.$$

Therefore we have the inequality  $I - \epsilon < J + \epsilon$  for any  $\epsilon > 0$ . This implies  $I \leq J$ .  $\square$

**Example 4.3.3.** If  $f(x)$  is integrable, then  $|f(x)|$  is integrable, and  $-|f(x)| \leq f(x) \leq |f(x)|$ . By Proposition 4.3.2, we have  $-\int_a^b |f(x)|dx \leq \int_a^b f(x)dx \leq \int_a^b |f(x)|dx$ . This is the same as

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx.$$

By  $\inf_{[a,b]} f \leq f(x) \leq \sup_{[a,b]} f$ , we also get

$$(b-a) \inf_{[a,b]} f = \int_a^b \left( \inf_{[a,b]} f \right) dx \leq \int_a^b f(x)dx \leq \int_a^b \left( \sup_{[a,b]} f \right) dx = (b-a) \sup_{[a,b]} f.$$

If  $f(x)$  is continuous, then this implies that the *average*  $\frac{1}{b-a} \int_a^b f(x)dx$  lies between the maximum and the minimum of  $f(x)$ , and by the Intermediate Value Theorem, we have

$$\int_a^b f(x)dx = f(c)(b-a) \text{ for some } c \in [a, b].$$

**Exercise 4.20.** Suppose  $f(x) \geq 0$  is a concave function on  $[a, b]$ . Then for any  $y \in [a, b]$ ,  $f(x)$  is bigger than the function obtained by connecting straight lines from  $(a, 0)$  to  $(y, f(y))$  and then to  $(b, 0)$ . Use this to prove that  $f(y) \leq \frac{2}{b-a} \int_a^b f(x)dx$ . Moreover, determine when the equality holds.

**Exercise 4.21.** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Suppose  $m \leq f' \leq M$  on  $(a, b)$  and denote  $\mu = \frac{f(b) - f(a)}{b - a}$ . By comparing  $f(x)$  with suitable piecewise linear functions, prove that

$$\left| \int_a^b f(x)dx - \frac{f(a) + f(b)}{2}(b-a) \right| \leq \frac{(M - \mu)(\mu - m)}{2(M - m)}(b-a)^2.$$

**Exercise 4.22 (First Integral Mean Value Theorem).** Suppose  $f(x)$  is continuous on  $[a, b]$  and  $g(x)$  is non-negative and integrable on  $[a, b]$ . Prove that there is  $c \in [a, b]$ , such that

$$\int_a^b g(x)f(x)dx = f(c) \int_a^b g(x)dx.$$

In fact, we can achieve this by  $c \in (a, b)$ .

**Exercise 4.23.** Suppose  $f(x)$  is integrable on  $[a, b]$ . Prove that

$$\left| f(c)(b-a) - \int_a^b f(x)dx \right| \leq \omega_{[a,b]}(f)(b-a) \text{ for any } c \in [a, b].$$

**Exercise 4.24 (Integral Continuity).** Suppose  $f(x)$  is a continuous function on an open interval containing  $[a, b]$ . Prove that  $\lim_{t \rightarrow 0} \int_a^b |f(x+t) - f(x)|dx = 0$ . We will see in Exercise 4.91 that the continuity assumption is not needed.



**Proposition 4.3.3.** *Suppose  $f(x)$  is a function on  $[a, c]$ , and  $b \in [a, c]$ . Then  $f(x)$  is integrable on  $[a, c]$  if and only if its restrictions on  $[a, b]$  and  $[b, c]$  are integrable. Moreover,*

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx.$$

The proposition reflects the intuition that, if a region is divided into non-overlapping parts, then the whole area is the sum of the areas of the parts.

*Proof.* The proof is based on the study of the relation of the Riemann sums of the function on  $[a, b]$ ,  $[b, c]$  and  $[a, c]$ . Let  $P$  be a partition of  $[a, c]$ .

If  $P$  contains  $b$  as a partition point, then  $P$  is obtained by combining a partition  $P'$  of  $[a, b]$  and a partition  $P''$  of  $[b, c]$ . For any choice of  $x_i^*$  for  $P$  and the same choice for  $P'$  and  $P''$ , we have  $S(P, f) = S(P', f) + S(P'', f)$ .

If  $P$  does not contain  $b$  as a partition point, then  $b \in (x_{k-1}, x_k)$  for some  $k$ , and the new partition  $\tilde{P} = P \cup \{b\}$  is still obtained by combining a partition  $P'$  of  $[a, b]$  and a partition  $P''$  of  $[b, c]$  together. For any choice of  $x_i^*$  for  $P$ , we keep all  $x_i^*$  with  $i \neq k$  and introduce  $x_k'^* \in [x_{k-1}, b]$ ,  $x_k''^* \in [b, x_k]$  for  $\tilde{P}$ . Then  $S(\tilde{P}, f) = S(P', f) + S(P'', f)$  as before, and

$$\begin{aligned} |S(P, f) - S(P', f) - S(P'', f)| &= |S(P, f) - S(\tilde{P}, f)| \\ &= |f(x_k^*)(x_k - x_{k-1}) - f(x_k'^*)(b - x_{k-1}) - f(x_k''^*)(x_k - b)| \\ &\leq 2 \left( \sup_{[x_{k-1}, x_k]} |f| \right) \|P\|. \end{aligned}$$

Suppose  $f$  is integrable on  $[a, b]$  and  $[b, c]$ . Then by Proposition 4.1.2,  $f(x)$  is bounded on the two intervals. Thus  $|f(x)| < B$  for some constant  $B$  and all  $x \in [a, c]$ . Denote  $I = \int_a^b f(x)dx$  and  $J = \int_b^c f(x)dx$ . For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that for partitions  $P'$  of  $[a, b]$  and  $P''$  of  $[b, c]$  satisfying  $\|P'\| < \delta$  and  $\|P''\| < \delta$ , we have  $|S(P', f) - I| < \epsilon$  and  $|S(P'', f) - J| < \epsilon$ . Then for any partition  $P$  of  $[a, c]$  satisfying  $\|P\| < \delta$ , we always have

$$\begin{aligned} |S(P, f) - I - J| &\leq |S(P, f) - S(P', f) - S(P'', f)| + |S(P', f) - I| + |S(P'', f) - J| \\ &< 2B\delta + 2\epsilon. \end{aligned}$$

This implies that  $f(x)$  is integrable on  $[a, c]$ , with  $\int_a^c f(x)dx = I + J$ .

It remains to show that the integrability on  $[a, c]$  implies the integrability on  $[a, b]$  and  $[b, c]$ . By the Cauchy criterion, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that for any partitions  $P$  and  $Q$  of  $[a, c]$  satisfying  $\|P\| < \delta$  and  $\|Q\| < \delta$ , we have  $|S(P, f) - S(Q, f)| < \epsilon$ . Now suppose  $P'$  and  $Q'$  are partitions of  $[a, b]$  satisfying  $\|P'\| < \delta$  and  $\|Q'\| < \delta$ . Let  $R$  be any partition of  $[b, c]$  satisfying  $\|R\| < \delta$ . By adding  $R$  to  $P'$  and  $Q'$ , we get partitions  $P$  and  $Q$  of  $[a, c]$  satisfying  $\|P\| < \delta$  and

$\|Q\| < \delta$ . Moreover, the choices of  $x_i^*$  (which may be different for  $P'$  and  $Q'$ ) may be extended by adding the same  $x_i^*$  for  $R$ . Then we get

$$\begin{aligned} |S(P', f) - S(Q', f)| &= |(S(P', f) + S(R, f)) - (S(Q', f) + S(R, f))| \\ &= |S(P, f) - S(Q, f)| < \epsilon. \end{aligned}$$

This proves the integrability of  $f$  on  $[a, b]$ . The proof of the integrability on  $[b, c]$  is similar.  $\square$

**Example 4.3.4.** A function  $f(x)$  on  $[a, b]$  is a *step function* if there is a partition  $P$  and constants  $c_i$ , such that  $f(x) = c_i$  on  $(x_{i-1}, x_i)$  (it does not matter what  $f(x_i)$  are). Then

$$\int_a^b f(x)dx = \sum \int_{x_{i-1}}^{x_i} f(x)dx = \sum \int_{x_{i-1}}^{x_i} c_i dx = \sum c_i(x_i - x_{i-1}).$$

**Example 4.3.5.** Suppose  $f(x) \geq 0$  is continuous on  $[a, b]$ . We claim that

$$\lim_{p \rightarrow +\infty} \left( \int_a^b f(x)^p dx \right)^{\frac{1}{p}} = \max_{[a, b]} f(x).$$

Let  $M = \max_{[a, b]} f(x)$ . We have  $M = f(x_0)$  for some  $x_0 \in [a, b]$ . Since  $f(x)$  is continuous at  $x_0$ , for any  $\epsilon > 0$ , we have  $f(x) > M - \epsilon$  on an interval  $[c, d]$  containing  $x_0$ . Then

$$\begin{aligned} M^p(b-a) &\geq \int_a^b f(x)^p dx = \left( \int_a^c + \int_c^d + \int_d^b \right) f(x)^p dx \\ &\geq \int_c^d f(x)^p dx \geq (M - \epsilon)^p(d - c). \end{aligned}$$

Therefore

$$M(b-a)^{\frac{1}{p}} \geq \left( \int_a^b f(x)^p dx \right)^{\frac{1}{p}} \geq (M - \epsilon)(d - c)^{\frac{1}{p}}.$$

As  $p \rightarrow +\infty$ , the left side converges to  $M$  and the right side converges to  $M - \epsilon$ . Therefore there is  $N$ , such that

$$\begin{aligned} p > N &\implies M + \epsilon > M(b-a)^{\frac{1}{p}}, (M - \epsilon)(d - c)^{\frac{1}{p}} > M - 2\epsilon \\ &\implies M + \epsilon > \left( \int_a^b f(x)^p dx \right)^{\frac{1}{p}} > M - 2\epsilon. \end{aligned}$$

This proves that  $\lim_{p \rightarrow +\infty} \left( \int_a^b f(x)^p dx \right)^{\frac{1}{p}} = M$ .

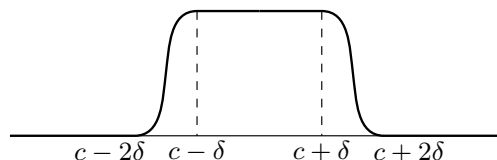
**Example 4.3.6 (Testing Function).** Suppose  $f(x)$  is continuous on  $[a, b]$ . If  $\int_a^b f(x)g(x)dx = 0$  for any continuous function  $g(x)$ , we prove that  $f(x) = 0$  everywhere.

Suppose  $f(c) > 0$  for some  $c \in [a, b]$ . By the continuity of  $f(x)$  at  $c$ , there is  $\delta > 0$ , such that  $f(x) > \frac{1}{2}f(c)$  on  $(c - 2\delta, c + 2\delta)$ . Then we construct a continuous function  $g(x)$

satisfying  $0 \leq g(x) \leq 1$ ,  $g(x) = 1$  on  $[c - \delta, c + \delta]$ , and  $g(x) = 0$  on  $[a, b] - (c - 2\delta, c + 2\delta)$ . See Figure 4.3.1. We have

$$\begin{aligned} \int_a^b f(x)g(x)dx &= \left( \int_a^{c-2\delta} + \int_{c-2\delta}^{c-\delta} + \int_{c-\delta}^{c+\delta} + \int_{c+\delta}^{c+2\delta} + \int_{c+2\delta}^b \right) f(x)g(x)dx \\ &= \left( \int_{c-2\delta}^{c-\delta} + \int_{c-\delta}^{c+\delta} + \int_{c+\delta}^{c+2\delta} \right) f(x)g(x)dx \\ &\geq \int_{c-\delta}^{c+\delta} f(x)g(x)dx = \int_{c-\delta}^{c+\delta} f(x)dx \geq f(c)\delta > 0. \end{aligned}$$

The first equality is due to Proposition 4.3.3. The second equality is due to  $g = 0$  on  $[a, c - 2\delta]$  and  $[c + 2\delta, b]$ . The third equality is due to  $g(x) = 1$  on  $[c - \delta, c + \delta]$ . The first inequality is due to  $f(x)g(x) \geq 0$  on  $[c - 2\delta, c - \delta]$  and  $[c + \delta, c + 2\delta]$ . The second inequality is due to  $f(x) > \frac{1}{2}f(c)$  on  $(c - 2\delta, c + 2\delta)$ . The contradiction at the end shows that  $f(x)$  cannot take positive value. The same argument shows that the function cannot take negative value. Therefore it is constantly zero.



**Figure 4.3.1.** Continuous (or even smooth) testing function  $g$ .

The function  $g(x)$  serves as a *testing function* for showing that a given function  $f(x)$  is constantly zero. This is a very useful technique. Note that with the help of the function in Example 3.4.9, the function in Figure 4.3.1 can be constructed to have derivatives of any order. In other words, we can use smooth functions as testing functions.

**Exercise 4.25.** Suppose  $f$  is continuous on  $[a, b]$  and  $f > 0$  on  $(a, b)$ . Suppose  $g$  is integrable on  $[a, b]$ . Prove that  $\lim_{p \rightarrow +\infty} \int_a^b g f^{\frac{1}{p}} dx = \int_a^b g dx$ .

**Exercise 4.26.** Suppose  $f$  is strictly increasing on  $[a, b]$  and  $f(a) \geq -1$ .

1. Prove that if  $f(b) \leq 1$ , then  $\lim_{p \rightarrow +\infty} \int_a^b f^p dx = 0$ .
2. Prove that if  $f(b) > 1$  and  $f$  is continuous at  $b$ , then  $\lim_{p \rightarrow +\infty} \int_a^b f^p dx = +\infty$ .

Extend the result to  $\lim_{p \rightarrow +\infty} \int_a^b f^p g dx$ , where  $g$  is non-negative and integrable on  $[a, b]$ .

**Exercise 4.27.** Suppose  $f$  is continuous on  $[a, b]$ . Prove that the following are equivalent.

1.  $f = 0$  on  $[a, b]$ .
2.  $\int_a^b |f| dx = 0$ .

3.  $\int_c^d f dx = 0$  for any  $[c, d] \subset [a, b]$ .
4.  $\int_a^b f g dx = 0$  for any smooth function  $g$ .

**Exercise 4.28.** Suppose  $f$  is continuous on  $[a, b]$ . Prove that  $\left| \int_a^b f dx \right| = \int_a^b |f| dx$  if and only if  $f$  does not change sign.

**Exercise 4.29.** Suppose  $f$  is continuous on  $[a, b]$  and satisfies  $\int_a^b f(x) dx = \int_a^b x f(x) dx = 0$ . Prove that  $f(c_1) = f(c_2) = 0$  at least two distinct points  $c_1, c_2 \in (a, b)$ . Extend the result to more points.

**Exercise 4.30.** Suppose  $f$  is a periodic integrable function of period  $T$ . Prove that  $\int_a^{a+T} f dx = \int_0^T f dx$ . Conversely, prove that if  $f$  is continuous and  $\int_a^{a+T} f dx$  is independent of  $a$ , then  $f$  is periodic of period  $T$ .

**Exercise 4.31.** Suppose  $f$  is a periodic integrable function of period  $T$ . Prove that

$$\lim_{b \rightarrow +\infty} \frac{1}{b} \int_a^b f dx = \frac{1}{T} \int_0^T f dx.$$

This says that the limit of the average on bigger and bigger intervals is the average on an interval of the period length.

**Exercise 4.32.** Suppose  $f$  is integrable on  $[a, b]$ . Prove that

$$\left| S(P, f) - \int_a^b f dx \right| \leq \omega(P, f).$$

**Exercise 4.33.** Suppose  $f$  satisfies the Lipschitz condition  $|f(x) - f(x')| \leq L|x - x'|$  on  $[a, b]$ . Prove that

$$\left| S(P, f) - \int_a^b f dx \right| \leq \frac{L}{2} \sum \Delta x_i^2.$$

This gives an estimate of how close the Riemann sum is to the actual integral.

**Exercise 4.34.** Suppose  $g$  is integrable on  $[a, b]$ . Prove that

$$\left| \sum f(x_i^*) \int_{x_{i-1}}^{x_i} g dx - \int_a^b f g dx \right| \leq \left( \sup_{[a, b]} |g| \right) \omega(P, f).$$

**Exercise 4.35.** Suppose  $f$  and  $g$  are integrable. Prove that

$$\lim_{\|P\| \rightarrow 0} \sum f(x_i^*) \int_{x_{i-1}}^{x_i} g dx = \int_a^b f g dx.$$

**Exercise 4.36.** Suppose  $f$  is integrable and satisfies  $\int_a^x f dx = 0$  for all  $x \in [a, b]$ . Prove that  $\int_a^b fg dx = 0$  for any integrable  $g$ .

**Exercise 4.37.** Suppose  $f$  and  $g$  are integrable on  $[a, b]$ . Prove that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that for any partition  $P$  satisfying  $\|P\| < \delta$  and choices  $x_i^*, x_i^{**} \in [x_{i-1}, x_i]$ , we have

$$\left| \sum f(x_i^*)g(x_i^{**})\Delta x_i - \int_a^b fg dx \right| < \epsilon.$$

**Exercise 4.38 (Riemann-Lebesgue Lemma).** Suppose  $f$  is a periodic function of period  $T$  and  $g$  is integrable on  $[a, b]$ . Prove that

$$\lim_{t \rightarrow \infty} \int_a^b f(tx)g(x)dx = \frac{1}{T} \int_0^T f(x)dx \int_a^b g(x)dx.$$

**Exercise 4.39.** Study Propositions 4.3.1, 4.3.2 and 4.3.3 for the lower and upper Darboux integrals.

The definition of the Riemann integral  $\int_a^b f(x)dx$  implicitly assumes  $a < b$ . If  $a > b$ , then we also define

$$\int_a^b f(x)dx = - \int_b^a f(x)dx.$$

Moreover, we define  $\int_a^a f(x)dx = 0$  (which can be considered as a special case of the original definition of the Riemann integral). Then the equality

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx$$

still holds for any order between  $a, b, c$ . Proposition 4.3.1 still holds for  $a \geq b$ , and the direction of the inequality in Proposition 4.3.2 needs to be reversed for  $a \geq b$ .

## 4.4 Fundamental Theorem of Calculus

The integration was originally considered as the inverse of the differentiation. Such connection between integration and differentiation is the *Fundamental Theorem of Calculus*.

### Newton-Leibniz Formula

**Theorem 4.4.1.** Suppose  $F(x)$  is continuous on  $[a, b]$ , differentiable on  $(a, b)$ , and  $F'(x)$  is integrable. Then  $F(x) = F(a) + \int_a^x F'(t)dt$ .

*Proof.* Let  $P$  be a partition of  $[a, b]$ . By the Mean Value Theorem, we can express  $F(b) - F(a)$  as a Riemann sum of  $F'(x)$  with suitable choice of  $x_i^*$

$$F(b) - F(a) = \sum (F(x_i) - F(x_{i-1})) = \sum F'(x_i^*)(x_i - x_{i-1}).$$

By the integrability of  $F'(x)$ , the right side converges to  $\int_a^b F'(x)dx$  as  $\|P\| \rightarrow 0$ .

Therefore  $F(b) - F(a) = \int_a^b F'(x)dx$ . □

The theorem tells us that, if  $f(x)$  is integrable and  $F(x)$  is an *antiderivative* of  $f$  (meaning  $F$  is differentiable and  $F'(x) = f(x)$ ), then we have the *Newton-Leibniz formula* for calculating the integral of  $f$

$$\int_a^b f(x)dx = F(b) - F(a).$$

This raises two questions.

1. Do all integrable functions have antiderivative?
2. Are the derivative of all differentiable functions integrable?

We will have examples showing that both answers are no. The following implies that continuous (stronger than integrable) functions have antiderivatives.

**Theorem 4.4.2.** *Suppose  $f(x)$  is integrable. Then  $F(x) = \int_a^x f(t)dt$  is a continuous function. Moreover, if  $f(x)$  is continuous at  $x_0$ , then  $F(x)$  is differentiable at  $x_0$ , with  $F'(x_0) = f(x_0)$ .*

*Proof.* By Proposition 4.1.2,  $f(x)$  is bounded by a constant  $B$ . Then by Example 4.3.3 and Proposition 4.3.3, we have

$$|F(x) - F(x_0)| = \left| \int_{x_0}^x f(t)dt \right| \leq B|x - x_0|.$$

This implies  $\lim_{x \rightarrow x_0} F(x) = F(x_0)$ .

Suppose we also know that  $f(x)$  is continuous at  $x_0$ . Then for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $|x - x_0| < \delta$  implies  $|f(x) - f(x_0)| < \epsilon$ . This further implies

$$\begin{aligned} |F(x) - F(x_0) - f(x_0)(x - x_0)| &= \left| \int_{x_0}^x f(t)dt - f(x_0)(x - x_0) \right| \\ &= \left| \int_{x_0}^x (f(t) - f(x_0))dt \right| \leq \epsilon|x - x_0|. \end{aligned}$$

Therefore  $F(x_0) + f(x_0)(x - x_0)$  is the linear approximation of  $F(x)$  at  $x_0$ . □

**Example 4.4.1.** The sign function  $\text{sign}(x)$  in Example 2.3.1 is integrable, with

$$\int_0^x \text{sign}(t) dt = \begin{cases} \int_0^x 1 dx = x, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ \int_0^x -1 dx = -x & \text{if } x < 0 \end{cases} = |x|.$$

If the sign function has an antiderivative  $F$ , then Theorem 4.4.1 implies  $F(x) = F(0) + |x|$ . Therefore  $F$  is not differentiable at  $x = 0$ . The non-differentiability confirms the necessity of the continuity at  $x_0$  in Theorem 4.4.2.

The example shows that an integrable function may not always have antiderivative. A generalisation of the example can be found in Exercise 4.41.

**Exercise 4.40.** If  $f(x)$  is not continuous at  $x_0$ , is it true that  $F(x) = \int_a^x f(t) dt$  is not differentiable at  $x_0$ ?

**Exercise 4.41.** Suppose  $f(x)$  is integrable and  $F(x) = \int_a^x f(t) dt$ . Prove that if  $f(x)$  has left limit at  $x_0$ , then  $F'_-(x_0) = f(x_0^-)$ . In particular, this shows that if  $f$  has different left and right limits at  $x_0$ , then  $F(x)$  is not differentiable at  $x_0$ .

**Exercise 4.42.** Suppose  $f(x)$  is integrable on  $[a, b]$ . Prove that  $\frac{1}{2} \int_a^b f(x) dx = \int_a^c f(x) dx$  for some  $c \in (a, b)$ . Can  $\frac{1}{2}$  be replaced by some other number?

**Example 4.4.2.** The function

$$F(x) = \begin{cases} x^2 \sin \frac{1}{x}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

is differentiable, and its derivative function

$$f(x) = \begin{cases} 2x \sin \frac{1}{x} - \cos \frac{1}{x}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

is not continuous but still integrable by Proposition 4.1.4 and Exercise 4.7. Therefore we still have the Newton-Leibniz formula  $\int_0^1 f(x) dx = F(1) - F(0) = \sin 1$ .

**Example 4.4.3.** The function

$$F(x) = \begin{cases} x^2 \sin \frac{1}{x^2}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

is differentiable, and its derivative function

$$f(x) = \begin{cases} 2x \sin \frac{1}{x^2} - \frac{2}{x} \cos \frac{1}{x^2}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

is not integrable on  $[0, 1]$  because it is not bounded.

The example appears to tell us that derivative functions are not necessarily integrable. However, such interpretation is misleading because the problem is not the failure of the integrability criterion, but that the integration is not defined for unbounded function. One natural way to extend the integration to unbounded functions is through *improper integral*, which is defined as the limit of the integration of the bounded part

$$\int_0^1 f(x)dx = \lim_{a \rightarrow 0^+} \int_a^1 f(x)dx = \lim_{a \rightarrow 0^+} (F(1) - F(a)) = F(1) - F(0) = \sin 1.$$

The example suggests that, by suitably extending the integration, derivative functions are likely to be always integrable. Moreover, the current example, Example 4.4.2, and Exercise 4.43 suggests that the Newton-Leibniz formula are likely always true in the extended integration theory.

**Exercise 4.43.** Prove that Theorem 4.4.1 still holds for *piecewise differentiable*. In other words, if  $F$  is continuous, differentiable at all but finitely many points, and  $F'$  is integrable, then the Newton-Leibniz formula remains true.

**Exercise 4.44.** Suppose  $F$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Suppose  $F(x) = F(a) + \int_a^x f(t)dt$  for an integrable  $f$  and all  $x \in (a, b)$ .

1. Prove that on any interval  $[c, d] \subset [a, b]$ , we have  $\inf_{[c, d]} f \leq F'(x) \leq \sup_{[c, d]} f$  for any  $x \in [c, d]$ . In other words, the derivative  $F'$  is bounded by the bounds of  $f$ .
2. Prove that  $F'$  is integrable on  $[a, b]$ .

Combined with Theorem 4.4.1, the second statement gives the necessary and sufficient condition for the integrability of  $F'$ .

**Exercise 4.45.** Explain that the function  $f$  in Exercise 4.44 is not necessarily equal to  $F'$ . Then use Exercise 4.36 to compare  $f$  and  $F'$ .

If the answers to both questions after Theorem 4.4.1 are affirmative, then we would have the perfect Fundamental Theorem. The examples and exercises above show that, although the Fundamental Theorem for Riemann integral is not perfect, the problem happens only at few places. They also suggest that, if the integration can be extended to tolerate “small defects”, then it is possible to get perfect Fundamental Theorem. This is realised by Theorem 12.3.5, in the setting of Lebesgue integral.

## Application of Fundamental Theorem

**Example 4.4.4.** We try to find continuous  $f(x)$  on  $[0, +\infty)$ , such that  $p \int_0^x tf(t)dt = x \int_0^x f(t)dt$ . The continuity means that we can take derivative on both sides and get  $pxf(x) = xf(x) + \int_0^x f(t)dt$ , or  $(p-1)xf(x) = \int_0^x f(t)dt$ .  
If  $p = 1$ , then  $\int_0^x f(t)dt = 0$  for all  $x \geq 0$ . This means  $f(x)$  is constantly zero.



If  $p \neq 1$ , then the differentiability of  $\int_0^x f(t)dt$  implies the differentiability of  $f(x) = \frac{1}{(p-1)x} \int_0^x f(t)dt$  for  $x > 0$ . By taking derivative of both sides of  $(p-1)xf(x) = \int_0^x f(t)dt$ , we get  $(p-1)(f(x) + xf'(x)) = f(x)$ , or  $qf(x) + xf'(x) = 0$ , where  $q = \frac{p-2}{p-1}$ . This means  $(x^q f(x))' = qx^{q-1}f(x) + x^q f'(x) = x^{q-1}(qf(x) + xf'(x)) = 0$ . Therefore we find that  $x^q f(x) = c$  is a constant, or  $f(x) = cx^{\frac{2-p}{p-1}}$ . We note that, in order for the integral on  $[0, x]$  to make sense,  $f(x)$  has to be bounded on bounded intervals. This means that we need to have  $\frac{2-p}{p-1} \geq 0$ , or  $1 < p \leq 2$ . In this case, substituting  $f(x) = cx^{\frac{2-p}{p-1}}$  into the original equation shows that it is indeed a solution.

We conclude that, if  $1 < p \leq 2$ , then the solutions are  $f(x) = cx^{\frac{2-p}{p-1}}$ . Otherwise, the only solution is  $f(x) = 0$ .

**Exercise 4.46.** Find continuous functions satisfying the equalities.

1.  $\int_0^x f(t)dt = \int_x^1 f(t)dt$  on  $[0, 1]$ .
2.  $p \int_1^x tf(t)dt = x \int_1^x f(t)dt$  on  $(0, +\infty)$ .
3.  $(f(x))^2 = 2 \int_0^x f(t)dt$  on  $(-\infty, +\infty)$ .

**Exercise 4.47.** Find continuous functions  $f(x)$  on  $(0, +\infty)$ , such that for all  $b > 0$ , the integral  $\int_a^{ab} f(x)dx$  is independent of  $a > 0$ .

Since integration is almost the converse of differentiation, the properties of differentiation should have integration counterparts. The counterpart of  $(f+g)' = f' + g'$  and  $(cf)' = cf'$  is Proposition 4.3.1. The counterpart of the Leibniz rule is the following.

**Theorem 4.4.3 (Integration by Parts).** *Suppose  $f(x), g(x)$  are differentiable, and  $f'(x), g'(x)$  are integrable. Then*

$$\int_a^b f(x)g'(x)dx + \int_a^b f'(x)g(x)dx = f(b)g(b) - f(a)g(a).$$

Since the current Fundamental Theorem is not perfect, the integration by parts in the theorem is not the best we can get. Theorem 4.5.3 gives the best version.

*Proof.* Since differentiability implies continuity (Proposition 3.1.3), and continuity implies integrability (Proposition 4.1.4), we know  $f, g$  are integrable. Then by Example 4.3.1,  $f'g', f'g$  are integrable, and  $(fg)' = f'g + fg'$  is integrable. By

Theorem 4.4.1, we get

$$\int_a^b f(x)g'(x)dx + \int_a^b f'(x)g(x)dx = \int_a^b (f(x)g(x))'dx = f(b)g(b) - f(a)g(a). \quad \square$$

The chain rule also has its integration counterpart.

**Theorem 4.4.4 (Change of Variable).** *Suppose  $\phi(x)$  is differentiable, with  $\phi'(x)$  integrable on  $[a, b]$ . Suppose  $f(y)$  is continuous on  $\phi([a, b])$ . Then*

$$\int_{\phi(a)}^{\phi(b)} f(y)dy = \int_a^b f(\phi(x))\phi'(x)dx.$$

Again the theorem is not the best change of variable formula we can get. Theorem 4.5.4 gives the best version. Moreover, Exercise 4.65 gives another version of change of variable, with more strict condition on  $\phi$  and less strict condition on  $f$ .

*Proof.* Applying Theorem 4.4.2 to the continuous  $f$ , we find that  $F(z) = \int_{\phi(a)}^z f(y)dy$  satisfies  $F'(z) = f(z)$ . Since  $\phi$  is differentiable, by the chain rule, we have

$$F(\phi(x))' = F'(\phi(x))\phi'(x) = f(\phi(x))\phi'(x).$$

Since  $f$  and  $\phi$  are continuous, the composition  $f(\phi(x))$  is continuous and therefore integrable. Moreover,  $\phi'(x)$  is assumed to be integrable. Therefore the product  $f(\phi(x))\phi'(x)$  is integrable, and we may apply Theorem 4.4.1 to get

$$F(\phi(b)) - F(\phi(a)) = \int_a^b f(\phi(x))\phi'(x)dx.$$

By the definition of  $F(z)$ , this is the equality in the theorem.  $\square$

**Exercise 4.48 (Jean Bernoulli<sup>22</sup>).** Suppose  $f(t)$  has continuous  $n$ -th order derivative on  $[0, x]$ . Prove that

$$\int_0^x f(t)dt = xf(x) - \frac{x^2}{2!}f'(x) + \cdots + (-1)^{n-1}\frac{x^n}{n!}f^{(n-1)}(x) + (-1)^n\frac{1}{n!}\int_0^x t^n f^{(n)}(t)dt.$$

**Exercise 4.49.** Suppose  $u(x)$  and  $v(x)$  have continuous  $n$ -th order derivatives on  $[a, b]$ . Prove that

$$\int_a^b uv^{(n)}dx = \left[uv^{(n-1)} - u'v^{(n-2)} + \cdots + (-1)^{n-1}u^{(n-1)}v\right]_{x=a}^{x=b} + (-1)^n \int_a^b u^{(n)}vdx.$$

Then apply the formula to  $\int_{x_0}^x (x-t)^n f^{(n+1)}(t)dt$  to prove the *integral form* of the remainder of the Taylor expansion

$$R_n(x) = \frac{1}{n!} \int_{x_0}^x (x-t)^n f^{(n+1)}(t)dt.$$

<sup>22</sup>Jean Bernoulli, born 1667 and died 1748 in Basel (Switzerland).

Exercise 4.50. Suppose  $f'(x)$  is integrable and  $\lim_{x \rightarrow +\infty} f'(x) = 0$ . Prove that

$$\lim_{b \rightarrow +\infty} \frac{1}{b} \int_a^b f(x) \sin x dx = 0.$$

Moreover, extend the result to high order derivatives.

Exercise 4.51. Suppose  $f(x)$  is continuous on  $[-a, a]$ . Prove that the following are equivalent.

1.  $f(x)$  is an odd function.
2.  $\int_{-b}^b f(x) dx = 0$  for any  $0 < b < a$ .
3.  $\int_{-a}^a f(x)g(x) dx = 0$  for any even continuous function  $g(x)$ .
4.  $\int_{-a}^a f(x)g(x) dx = 0$  for any even integrable function  $g(x)$ .

Exercise 4.52. Suppose  $f(x)$  is integrable on  $[0, 1]$  and is continuous at 0. Prove that

$$\lim_{h \rightarrow 0^+} \int_0^1 \frac{h}{h^2 + x^2} f(x) dx = \frac{\pi}{2} f(0).$$

Exercise 4.53. Suppose  $f(x)$  is integrable on an open interval containing  $[a, b]$  and is continuous at  $a$  and  $b$ . Prove that

$$\lim_{h \rightarrow 0} \int_a^b \frac{f(x+h) - f(x)}{h} dx = f(b) - f(a).$$

The result should be compared with the equality

$$\int_a^b \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} dx = \int_a^b f'(x) dx = f(b) - f(a),$$

which by Theorem 4.4.1 holds when  $f(x)$  is differentiable and  $f'(x)$  is integrable.

## 4.5 Riemann-Stieltjes Integration

### Riemann-Stieltjes Sum

Let  $f$  and  $\alpha$  be functions on a bounded interval  $[a, b]$ . The *Riemann-Stieltjes*<sup>23</sup> sum of  $f$  with respect to  $\alpha$  is

$$S(P, f, \alpha) = \sum_{i=1}^n f(x_i^*)(\alpha(x_i) - \alpha(x_{i-1})) = \sum_{i=1}^n f(x_i^*) \Delta \alpha_i.$$

<sup>23</sup>Thomas Jan Stieltjes, born 1856 in Zwolle (Netherlands), died 1894 in Toulouse (France). He is often called the father of the analytic theory of continued fractions and is best remembered for his integral.

We say that  $f$  has *Riemann-Stieltjes integral*  $I$  and denote  $\int_a^b f d\alpha = I$ , if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|P\| < \delta \implies |S(P, f, \alpha) - I| < \epsilon.$$

When  $\alpha(x) = x$ , we get the Riemann integral.

**Example 4.5.1.** Suppose  $f(x) = c$  is a constant. Then  $S(P, f, \alpha) = c(\alpha(b) - \alpha(a))$ . Therefore  $\int_a^b c d\alpha = c(\alpha(b) - \alpha(a))$ .

**Example 4.5.2.** Suppose  $\alpha(x) = \alpha_0$  is a constant. Then  $\Delta\alpha_i = 0$ , and therefore  $\int_a^b f d\alpha_0 = 0$  for any  $f$ . Since  $f$  can be arbitrary, we see that Proposition 4.1.2 cannot be extended without additional conditions.

**Example 4.5.3.** Suppose the Dirichlet function  $D(x)$  is Riemann-Stieltjes integrable with respect to  $\alpha$  on  $[a, b]$ . We claim that  $\alpha$  must be a constant.

Let  $c, d \in [a, b]$ ,  $c < d$ . Let  $P$  be any partition with  $c$  and  $d$  as partition points. If we choose all  $x_i^*$  to be irrational, then we get  $S(P, D, \alpha) = 0$ . If we choose  $x_i^*$  to be irrational whenever  $[x_{i-1}, x_i] \not\subset [c, d]$  and choose  $x_i^*$  to be rational whenever  $[x_{i-1}, x_i] \subset [c, d]$ , then we get  $S(P, D, \alpha) = \alpha(d) - \alpha(c)$ . Since the two choices should converge to the same limit, we conclude that  $\alpha(d) - \alpha(c) = 0$ . Since  $c, d$  are arbitrary, we conclude that  $\alpha$  is a constant.

The example shows that it is possible for a nonzero function  $f$  to satisfy  $\int_a^b f d\alpha = 0$  whenever the Riemann-Stieltjes integral makes sense.

**Exercise 4.54.** Prove that the only function that is Riemann-Stieltjes integrable with respect to the Dirichlet function is the constant function. In particular, this shows that  $\int_a^b f d\alpha = 0$  for any function  $f$  that is Riemann-Stieltjes integrable with respect to  $\alpha$  does not necessarily imply that  $\alpha$  is a constant.

**Exercise 4.55.** Suppose  $c \in (a, b)$ , the three numbers  $\alpha_-, \alpha_0, \alpha_+$  are not all equal, and

$$\alpha(x) = \begin{cases} \alpha_-, & \text{if } x < c, \\ \alpha_0, & \text{if } x = c, \\ \alpha_+, & \text{if } x > c. \end{cases}$$

Prove that  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  if and only if  $f$  is continuous at  $c$ . Moreover, we have  $\int_a^b f d\alpha = f(c)(\alpha_+ - \alpha_-)$ .

**Exercise 4.56.** Find suitable  $\alpha$  on  $[0, 2]$ , such that  $\int_0^2 f d\alpha = f(0) + f(1) + f(2)$  for any continuous  $f$  on  $[0, 2]$ .

**Proposition 4.5.1.** *Suppose  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  on  $[a, b]$ . Then at any  $c \in [a, b]$ , either  $f$  or  $\alpha$  is continuous at  $c$ .*

*Proof.* The Cauchy criterion for the convergence of Riemann-Stieltjes sum is that, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|P\| < \delta, \|P'\| < \delta \implies |S(P, f, \alpha) - S(P', f, \alpha)| < \epsilon.$$

For  $x \leq u \leq v \leq y$  satisfying  $|x - y| < \delta$ , we choose  $P$  to be a partition with  $[x, y]$  as a partition interval and with  $u \in [x, y]$  as a sample point. We also choose  $P'$  to be the same partition, with the same sample points except  $u$  is replaced by  $v$ . Then we get

$$S(P, f, \alpha) - S(P', f, \alpha) = (f(u) - f(v))(\alpha(y) - \alpha(x)),$$

and the Cauchy criterion specialises to

$$x \leq u \leq v \leq y, |x - y| < \delta \implies |f(u) - f(v)| |\alpha(x) - \alpha(y)| < \epsilon.$$

Now we study what happens at  $c \in [a, b]$ . If  $\alpha$  is not continuous at  $c$ , then either  $\lim_{x \rightarrow c} \alpha(x)$  diverges, or the limit converges but is not equal to  $\alpha(c)$ . We want to show that  $f$  is continuous at  $c$  in either case.

Suppose  $\lim_{x \rightarrow c} \alpha(x)$  diverges. By Exercise 2.14, there is  $B > 0$ , such that for the  $\delta > 0$  above, we can find  $x, y$  satisfying  $c - \frac{\delta}{2} < x < c < y < c + \frac{\delta}{2}$  and  $|\alpha(x) - \alpha(y)| \geq B$ . Note that  $B$  depends only on  $\alpha$  and  $c$ , and is independent of  $\delta$  and  $\epsilon$ . Since  $|x - y| < \delta$ , we get

$$u, v \in [x, y] \implies |f(u) - f(v)| < \frac{\epsilon}{|\alpha(x) - \alpha(y)|} < \frac{\epsilon}{B}.$$

By  $x < c < y$ , we can find  $\delta'$ , such that  $x < c - \delta' < c < c + \delta' < y$ . Then by taking  $v = c$ , the implication above further implies

$$|u - c| < \delta' \implies |f(u) - f(c)| < \frac{\epsilon}{B}.$$

This verifies the continuity of  $f$  at  $c$ .

Suppose  $\lim_{x \rightarrow c} \alpha(x)$  converges but is not equal to  $\alpha(c)$ . By  $\lim_{x \rightarrow c^+} \alpha(x) \neq \alpha(c)$ , there is  $B > 0$ , such that for the  $\delta > 0$  above, we can find  $y$  satisfying  $c < y < c + \delta$  and  $|\alpha(c) - \alpha(y)| \geq B$ . Since  $|y - c| < \delta$ , we get

$$u, v \in [c, y] \implies |f(u) - f(v)| < \frac{\epsilon}{|\alpha(c) - \alpha(y)|} < \frac{\epsilon}{B}.$$

Let  $\delta' = y - c$ . Then by taking  $v = c$ , the implication above further implies

$$0 \leq u - c < \delta' \implies |f(u) - f(c)| < \frac{\epsilon}{B}.$$

This verifies the right continuity of  $f$  at  $c$ . By the same reason,  $f$  is also left continuous at  $c$ .  $\square$

## Property of Riemann-Stieltjes Integral

Many properties of the Riemann integral can be extended to Riemann-Stieltjes integral. Sometimes one needs to be careful with the continuity condition from Proposition 4.5.1. When it comes to inequality, we may also need  $\alpha$  to be monotone.

**Exercise 4.57.** Suppose  $f$  and  $g$  are Riemann-Stieltjes integrable with respect to  $\alpha$ . Prove that  $f + g$  and  $cf$  are Riemann-Stieltjes integrable with respect to  $\alpha$  and

$$\int_a^b (f + g)d\alpha = \int_a^b f d\alpha + \int_a^b g d\alpha, \quad \int_a^b c f d\alpha = c \int_a^b f d\alpha.$$

This extends Proposition 4.3.1.

**Exercise 4.58.** Suppose  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  and  $\beta$ . Suppose  $c$  is a constant. Prove that  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha + \beta$  and  $c\alpha$ , and

$$\int_a^b f d(\alpha + \beta) = \int_a^b f d\alpha + \int_a^b f d\beta, \quad \int_a^b f d(c\alpha) = c \int_a^b f d\alpha.$$

**Exercise 4.59.** Suppose  $f$  and  $\alpha$  are functions on  $[a, c]$  and  $b \in (a, c)$ .

1. Prove that if  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  on  $[a, c]$ , then  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  on  $[a, b]$  and  $[b, c]$ , and

$$\int_a^c f d\alpha = \int_a^b f d\alpha + \int_b^c f d\alpha.$$

2. Suppose  $f$  and  $\alpha$  are bounded, and either  $f$  or  $\alpha$  is continuous at  $c$ . Prove that if  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  on  $[a, b]$  and  $[b, c]$ , then  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  on  $[a, c]$ .
3. Construct bounded functions  $f$  and  $\alpha$ , such that  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  on  $[a, b]$  and  $[b, c]$ , but  $f$  is left continuous and not right continuous at  $c$ , while  $\alpha$  is right continuous and not left continuous at  $c$ . By Exercise 4.55,  $f$  is not Riemann-Stieltjes integrable with respect to  $\alpha$  on  $[a, c]$ .

This shows that Proposition 4.3.3 may be extended as long as either  $f$  or  $\alpha$  is continuous at the breaking point  $c$ .

**Exercise 4.60.** Suppose  $f$  and  $g$  are Riemann-Stieltjes integrable with respect to an increasing  $\alpha$  on  $[a, b]$ . Prove that

$$f \leq g \implies \int_a^b f d\alpha \leq \int_a^b g d\alpha.$$

Moreover, if  $\alpha$  is strictly increasing and  $f$  and  $g$  are continuous, then the equality holds if and only if  $f = g$ . This shows that Proposition 4.3.2 may be extended as long as we assume that  $\alpha$  is increasing.

**Exercise 4.61.** Suppose  $f$  is Riemann-Stieltjes integrable with respect to an increasing  $\alpha$  on  $[a, b]$ . Prove that

$$\begin{aligned} \left| \int_a^b f d\alpha \right| &\leq \int_a^b |f| d\alpha, \\ (\alpha(b) - \alpha(a)) \inf_{[a,b]} f &\leq \int_a^b f d\alpha \leq (\alpha(b) - \alpha(a)) \sup_{[a,b]} f, \\ \left| f(c)(\alpha(b) - \alpha(a)) - \int_a^b f d\alpha \right| &\leq \omega_{[a,b]}(f)(\alpha(b) - \alpha(a)) \text{ for } c \in [a, b], \\ \left| S(P, f, \alpha) - \int_a^b f d\alpha \right| &\leq \sum \omega_{[x_{i-1}, x_i]}(f) \Delta\alpha_i. \end{aligned}$$

Moreover, extend the first Integral Mean Value Theorem in Exercise 4.22 to the Riemann-Stieltjes integral.

**Exercise 4.62.** Suppose  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ , and  $F(x) = \int_a^x f d\alpha$ . Prove that if  $\alpha$  is monotone and  $f$  is continuous at  $x_0$ , then for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|\Delta x| = |x - x_0| < \delta \implies |F(x) - F(x_0) - f(x_0)\Delta\alpha| \leq \epsilon |\Delta\alpha|.$$

This shows that the Fundamental Theorem of Calculus may be extended as long as  $\alpha$  is monotone.

The following is the relation between Riemann-Stieltjes integral and the ordinary Riemann integral. An extension is given in Exercise 4.78. A much more general extension is given by Example 12.4.1.

**Theorem 4.5.2.** Suppose  $f$  is bounded,  $g$  is Riemann integrable, and  $\alpha(x) = \alpha(a) + \int_a^x g(t)dt$ . Then  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  if and only if  $fg$  is Riemann integrable. Moreover,

$$\int_a^b f d\alpha = \int_a^b fg dx.$$

For the special case  $\alpha$  is differentiable with Riemann integrable  $\alpha'$ , by Theorem 4.4.1, the formula becomes

$$\int_a^b f d\alpha = \int_a^b f \alpha' dx.$$

This is consistent with the notation  $d\alpha = \alpha' dx$  for the differential.

*Proof.* Suppose  $|f| < B$  for a constant  $B$ . Since  $g$  is Riemann integrable, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$  implies  $\sum \omega_{[x_{i-1}, x_i]}(g) \Delta x_i < \epsilon$ . For the

same choice of  $P$  and  $x_i^*$ , the Riemann-Stieltjes sum of  $f$  with respect to  $\alpha$  is

$$S(P, f, \alpha) = \sum f(x_i^*)(\alpha(x_i) - \alpha(x_{i-1})) = \sum f(x_i^*) \int_{x_{i-1}}^{x_i} g(t) dt.$$

The Riemann sum of  $fg$  is

$$S(P, fg) = \sum f(x_i^*)g(x_i^*)\Delta x_i.$$

When  $\|P\| < \delta$ , we have (the second inequality uses Exercise 4.23)

$$\begin{aligned} |S(P, f, \alpha) - S(P, fg)| &\leq \sum |f(x_i^*)| \left| \int_{x_{i-1}}^{x_i} g(t) dt - g(x_i^*)\Delta x_i \right| \\ &\leq \sum B\omega_{[x_{i-1}, x_i]}(g)\Delta x_i \leq B\epsilon. \end{aligned}$$

This implies that the Riemann-Stieltjes sum of  $f$  with respect to  $\alpha$  converges if and only if the Riemann sum of  $fg$  converges. Moreover, the two limits are the same.  $\square$

The following shows that, in a Riemann-Stieltjes integral  $\int_a^b f d\alpha$ , the roles of  $f$  and  $\alpha$  may be exchanged.

**Theorem 4.5.3** (Integration by Parts). *A function  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  if and only if  $\alpha$  is Riemann-Stieltjes integrable with respect to  $f$ . Moreover,*

$$\int_a^b f d\alpha + \int_a^b \alpha df = f(b)\alpha(b) - f(a)\alpha(a).$$

*Proof.* For a partition

$$P: a = x_0 < x_1 < x_2 < \cdots < x_n = b,$$

and sample points  $x_i^* \in [x_{i-1}, x_i]$ , we have the Riemann-Stieltjes sum of  $f$  with respect to  $\alpha$

$$S(P, f, \alpha) = \sum_{i=1}^n f(x_i^*)(\alpha(x_i) - \alpha(x_{i-1})).$$

We use the sample points to form

$$Q: a = x_0^* \leq x_1^* \leq x_2^* \leq \cdots \leq x_n^* \leq x_{n+1}^* = b,$$

which is almost a partition except some possible repetition among partition points. Moreover, if we use the partition points  $x_{i-1} \in [x_{i-1}^*, x_i^*]$ ,  $1 \leq i \leq n+1$ , of  $P$  as the sample points for  $Q$ , then the Riemann-Stieltjes sum of  $\alpha$  with respect to  $f$  is

$$S(Q, \alpha, f) = \sum_{i=1}^{n+1} \alpha(x_{i-1})(f(x_i^*) - f(x_{i-1}^*)).$$



Note that in case a repetition  $x_i^* = x_{i-1}^*$  happens, the corresponding term in the sum simply vanishes, so that  $S(Q, \alpha, f)$  is the same as the Riemann-Stieltjes sum after removing all the repetitions.

Now we add the two Riemann-Stieltjes sums together (see Figure 4.5.1 for the geometric meaning)

$$\begin{aligned}
& S(P, f, \alpha) + S(Q, \alpha, f) \\
&= \sum_{i=1}^n f(x_i^*)(\alpha(x_i) - \alpha(x_{i-1})) + \sum_{i=1}^{n+1} \alpha(x_{i-1})(f(x_i^*) - f(x_{i-1}^*)) \\
&= \sum_{i=1}^n f(x_i^*)\alpha(x_i) - \sum_{i=1}^n f(x_i^*)\alpha(x_{i-1}) + \sum_{i=1}^{n+1} \alpha(x_{i-1})f(x_i^*) - \sum_{i=1}^{n+1} \alpha(x_{i-1})f(x_{i-1}^*) \\
&= \sum_{i=1}^{n+1} \alpha(x_{i-1})f(x_i^*) - \sum_{i=1}^n f(x_i^*)\alpha(x_{i-1}) + \sum_{i=1}^n f(x_i^*)\alpha(x_i) - \sum_{i=1}^{n+1} \alpha(x_{i-1})f(x_{i-1}^*) \\
&= f(x_{n+1}^*)\alpha(x_n) - f(x_0)\alpha(x_0^*) = f(b)\alpha(b) - f(a)\alpha(a).
\end{aligned}$$

The limit of the equality gives us the integration by parts. Specifically, if  $\alpha$  is Riemann-Stieltjes integrable with respect to  $f$ , with  $I = \int_a^b \alpha df$ , then for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|Q\| < \delta \implies |S(Q, \alpha, f) - I| < \epsilon.$$

Since  $x_i^* - x_{i-1}^* \leq x_i - x_{i-2} \leq 2\|P\|$ , we have  $\|Q\| \leq 2\|P\|$ . Then  $\|P\| < \frac{\delta}{2}$  implies  $\|Q\| < \delta$ , which further implies

$$|S(P, f, \alpha) - (f(b)\alpha(b) - f(a)\alpha(a) - I)| = |S(Q, \alpha, f) - I| < \epsilon.$$

This proves that  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ , and the equality in the theorem holds.  $\square$

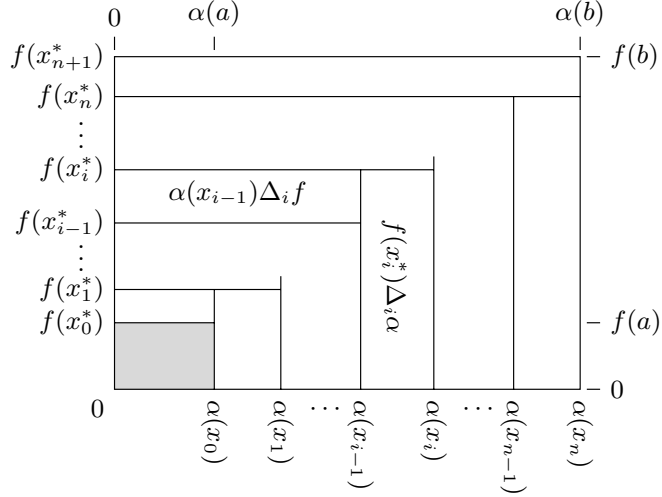
**Theorem 4.5.4 (Change of Variable).** *Suppose  $\phi$  is increasing and continuous on  $[a, b]$ . Suppose  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  on  $[\phi(a), \phi(b)]$ . Then  $f \circ \phi$  is Riemann-Stieltjes integrable with respect to  $\alpha \circ \phi$  on  $[a, b]$ . Moreover,*

$$\int_{\phi(a)}^{\phi(b)} f d\alpha = \int_a^b (f \circ \phi) d(\alpha \circ \phi).$$

*Proof.* Let  $P$  be a partition of  $[a, b]$ . Since  $\phi$  is increasing, we have a partition

$$\phi(P): \phi(a) = \phi(x_0) \leq \phi(x_1) \leq \phi(x_2) \leq \cdots \leq \phi(x_n) = \phi(b)$$

of  $[\phi(a), \phi(b)]$ . As explained in the proof of Theorem 4.5.3, we do not need to be worried about the repetition in partition points.



**Figure 4.5.1.** The sum of vertical and horizontal strips is  $f(b)\alpha(b) - f(a)\alpha(a)$ .

Choose  $x_i^*$  for  $P$  and choose the corresponding  $\phi(x_i^*)$  for  $\phi(P)$ . Then the Riemann-Stieltjes sum

$$S(P, f \circ \phi, \alpha \circ \phi) = \sum f(\phi(x_i^*))(\alpha(\phi(x_i)) - \alpha(\phi(x_{i-1}))) = S(\phi(P), f, \alpha).$$

Since  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ , for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|Q\| < \delta \implies \left| S(Q, f, \alpha) - \int_{\phi(a)}^{\phi(b)} f d\alpha \right| < \epsilon.$$

The continuity of  $\phi$  implies the uniform continuity. Therefore there is  $\delta' > 0$ , such that  $\|P\| < \delta'$  implies  $\|\phi(P)\| < \delta$ . Then  $\|P\| < \delta'$  implies

$$\left| S(P, f \circ \phi, \alpha \circ \phi) - \int_{\phi(a)}^{\phi(b)} f d\alpha \right| = \left| S(\phi(P), f, \alpha) - \int_{\phi(a)}^{\phi(b)} f d\alpha \right| < \epsilon.$$

This proves that  $f \circ \phi$  is Riemann-Stieltjes integrable with respect to  $\alpha \circ \phi$  and the equality in the theorem holds.  $\square$

**Exercise 4.63.** Prove the following version of the integration by parts: Suppose  $\phi(x)$ ,  $\psi(x)$  are integrable on  $[a, b]$ , and  $f(x) = f(a) + \int_a^x \phi(t)dt$ ,  $g(x) = g(a) + \int_a^x \psi(t)dt$ . Then

$$\int_a^b f(x)\psi(x)dx + \int_a^b \phi(x)g(x)dx = f(b)g(b) - f(a)g(a).$$

**Exercise 4.64.** What happens to Theorem 4.5.4 if  $\phi$  is decreasing? What if  $\phi(x)$  is not assumed to be continuous?

**Exercise 4.65.** Prove the following version of the change of variable: Suppose  $\phi(x) = \phi(a) + \int_a^x g(t)dt$  for an integrable  $g \geq 0$  on  $[a, b]$ . Suppose  $f(y)$  is integrable on  $[\phi(a), \phi(b)]$ . Then

$$\int_{\phi(a)}^{\phi(b)} f(y)dy = \int_a^b f(\phi(x))g(x)dx.$$

**Exercise 4.66 (Young's Inequality).** Suppose  $f$  and  $g$  are increasing functions satisfying  $g(f(x)) = x$  and  $f(0) = 0$ . Prove that if  $f$  is continuous, then for any  $a, b > 0$ , we have

$$\int_0^a f(x)dx + \int_0^b g(y)dy \geq ab.$$

Show that Young's inequality in Exercise 3.74 is a special case of this.

## 4.6 Bounded Variation Function

### Variation of Function

We want to extend the criterion for Riemann integrability (Theorem 4.1.3) to the Riemann-Stieltjes integral. The technical tool for Theorem 4.1.3 is the estimation such as (4.1.3). For monotone  $\alpha$ , it is straightforward to extend the estimation to

$$|f(c)(\alpha(b) - \alpha(a)) - S(P, f, \alpha)| \leq \omega_{[a,b]}(f)|\alpha(b) - \alpha(a)|,$$

and further get the straightforward extension of Theorem 4.1.3.

For general  $\alpha$ , however, the straightforward extension is wrong. So we introduce the *variation* of  $\alpha$  with respect to a partition  $P$

$$V_P(\alpha) = |\alpha(x_1) - \alpha(x_0)| + |\alpha(x_2) - \alpha(x_1)| + \cdots + |\alpha(x_n) - \alpha(x_{n-1})| = \sum |\Delta\alpha_i|.$$

If a partition  $Q$  refines  $P$ , then we clearly have  $V_Q(\alpha) \geq V_P(\alpha)$ . So we define the *variation* of  $\alpha$  on an interval

$$V_{[a,b]}(\alpha) = \sup\{V_P(\alpha) : P \text{ is a partition of } [a, b]\}.$$

We say that  $\alpha$  has *bounded variation* if the quantity is finite.

If  $|f| \leq B$ , then we have

$$|S(P, f, \alpha)| \leq \sum |f(x_i^*)||\Delta\alpha_i| \leq B \sum |\Delta\alpha_i| = BV_P(\alpha).$$

For any  $c \in [a, b]$ , we have  $|f(c) - f(x)| \leq \omega_{[a,b]}(f)$ . By the inequality above, we get

$$|f(c)(\alpha(b) - \alpha(a)) - S(P, f, \alpha)| = |S(P, f(c) - f(x), \alpha)| \leq \omega_{[a,b]}(f)V_P(\alpha), \quad (4.6.1)$$

**Proposition 4.6.1.** Suppose  $\alpha$  and  $\beta$  are bounded variation functions on  $[a, b]$ . Then  $\alpha + \beta$  and  $c\alpha$  are bounded variation functions, and

$$\begin{aligned} V_{[a,b]}(\alpha) &\geq |\alpha(b) - \alpha(a)|, \\ V_{[a,b]}(\alpha) &= V_{[a,c]}(\alpha) + V_{[c,b]}(\alpha) \text{ for } c \in (a, b), \\ V_{[a,b]}(\alpha + \beta) &\leq V_{[a,b]}(\alpha) + V_{[a,b]}(\beta), \\ V_{[a,b]}(c\alpha) &= |c|V_{[a,b]}(\alpha). \end{aligned}$$

Moreover  $V_{[a,b]}(\alpha) = |\alpha(b) - \alpha(a)|$  if and only if  $\alpha$  is monotone.

*Proof.* The first follows from  $V_P(\alpha) \geq |\alpha(b) - \alpha(a)|$  for any partition  $P$ .

For the second, we note that, since more refined partition gives bigger variation,  $V_{[a,b]}(\alpha)$  is also the supremum of  $V_P(\alpha)$  for those  $P$  with  $c$  as a partition point. Such  $P$  is the union of a partition  $P'$  of  $[a, c]$  and a partition  $P''$  of  $[c, b]$ , and we have  $V_P(\alpha) = V_{P'}(\alpha) + V_{P''}(\alpha)$ . By taking the supremum on both sides of the equality, we get  $V_{[a,b]}(\alpha) = V_{[a,c]}(\alpha) + V_{[c,b]}(\alpha)$ .

The third follows from  $V_P(\alpha + \beta) \leq V_P(\alpha) + V_P(\beta)$ . The fourth follows from  $V_P(c\alpha) = |c|V_P(\alpha)$ .

Finally, if  $\alpha$  is monotone, then  $\Delta\alpha_i$  have the same sign, and we get

$$V_{[a,b]}(\alpha) = \sum |\Delta\alpha_i| = \left| \sum \Delta\alpha_i \right| = |\alpha(b) - \alpha(a)|.$$

Conversely, suppose  $V_{[a,b]}(\alpha) = |\alpha(b) - \alpha(a)|$ . Then for any  $a \leq x < y \leq b$ , we take  $P$  to consist of  $a, x, y, b$  and get

$$|\alpha(b) - \alpha(a)| = V_{[a,b]}(\alpha) \geq V_P(\alpha) = |\alpha(x) - \alpha(a)| + |\alpha(y) - \alpha(x)| + |\alpha(b) - \alpha(y)|.$$

This implies that  $\alpha(y) - \alpha(x)$ ,  $\alpha(x) - \alpha(a)$ ,  $\alpha(b) - \alpha(y)$  always have the same sign. Therefore  $\alpha(y) - \alpha(x)$  and  $\alpha(b) - \alpha(a)$  also have the same sign for any  $x < y$ . This means that  $\alpha$  is monotone.  $\square$

**Example 4.6.1.** Suppose  $\alpha\left(\frac{1}{n}\right) = a_n$  and  $\alpha(x) = 0$  otherwise. By choosing the partition points  $x_i$  of  $[0, 1]$  to be  $0, \frac{1}{N}, c_N, \frac{1}{N-1}, c_{N-1}, \dots, \frac{1}{2}, c_2, 1$ , where  $c_i$  are not of the form  $\frac{1}{n}$ , we get  $V_P(\alpha) = 2 \sum_{i=1}^N |a_i| - |a_1|$ . Therefore a necessary condition for  $\alpha$  to have bounded variation is that any partial sum  $\sum_{i=1}^N |a_i|$  is bounded, which means that the series  $\sum a_n$  absolutely converges.

On the other hand, for any partition  $P$  of  $[0, 1]$ ,  $V_P(\alpha)$  is a sum of certain  $|a_n|$  and  $|a_n - a_{n-1}|$ . Moreover, each  $a_n$  either does not appear in the sum or appears twice in the sum, with the only exception that  $a_1$  always appears once. Therefore the sum is never more than the sum  $2 \sum_{n=1}^{\infty} |a_n| - |a_1|$ . Combined with the earlier discussion on the special choice of  $P$ , we conclude that  $V_{[0,1]}(\alpha) = 2 \sum_{i=1}^{\infty} |a_i| - |a_1|$ .

**Example 4.6.2.** Suppose  $g$  is Riemann integrable and  $\alpha(x) = \int_a^x g(t)dt$ . Then by Exercise 4.23,

$$\begin{aligned} |S(P, |g|) - V_P(\alpha)| &\leq \left| \sum |g(x_i^*)| \Delta x_i - \sum \left| \int_{x_{i-1}}^{x_i} g(t)dt \right| \right| \\ &\leq \sum \left| g(x_i^*) \Delta x_i - \int_{x_{i-1}}^{x_i} g(t)dt \right| \\ &\leq \sum \omega_{[x_{i-1}, x_i]}(g) \Delta x_i. \end{aligned}$$

By the Riemann integrability criterion, the right side can be arbitrarily small. Therefore

$\alpha$  has bounded variation, and

$$V_{[a,b]}(\alpha) = \int_a^b |g(t)| dt.$$

**Exercise 4.67.** Prove that any Lipschitz function has bounded variation.

**Exercise 4.68.** Prove that any bounded variation function is Riemann integrable.

**Exercise 4.69.** Suppose  $x_n$  is a non-repeating sequence in  $(a, b)$ . Suppose  $\alpha(x)$  is a function on  $[a, b]$  satisfying  $\alpha(x) = 0$  if  $x$  is not any  $x_n$ . Prove that  $V_{[a,b]} \alpha(x) = 2 \sum |\alpha(x_n)|$ . In particular, the function has bounded variation if and only if  $\sum \alpha(x_n)$  absolutely converges.

**Exercise 4.70.** Suppose  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ . Extend the inequalities in Exercise 4.61

$$\begin{aligned} \left| \int_a^b f d\alpha \right| &\leq \left( \sup_{[a,b]} |f| \right) V_{[a,b]}(\alpha), \\ \left| f(c)(\alpha(b) - \alpha(a)) - \int_a^b f d\alpha \right| &\leq \omega_{[a,b]}(f) V_{[a,b]}(\alpha) \text{ for } c \in [a, b], \\ \left| S(P, f, \alpha) - \int_a^b f d\alpha \right| &\leq \sum \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha). \end{aligned}$$

**Exercise 4.71.** Suppose  $F(x) = \int_a^x f d\alpha$  and  $\alpha$  has bounded variation. Prove that if  $f$  is continuous at  $x_0$ , then for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|\Delta x| = |x - x_0| < \delta \implies |F(x) - F(x_0) - f(x_0)\Delta\alpha| \leq \epsilon V_{[x_0, x]}(\alpha).$$

This extends Exercise 4.62.

**Exercise 4.72.** Suppose  $F(x) = \int_a^x f d\alpha$  and  $\alpha$  has bounded variation. Prove that if  $\alpha$  is not continuous at  $x_0$  and  $f(x_0) \neq 0$ , then  $F$  is not continuous at  $x_0$ .

## Decomposition of Bounded Variation Function

A bounded variation function  $\alpha$  on  $[a, b]$  induces a *variation function*

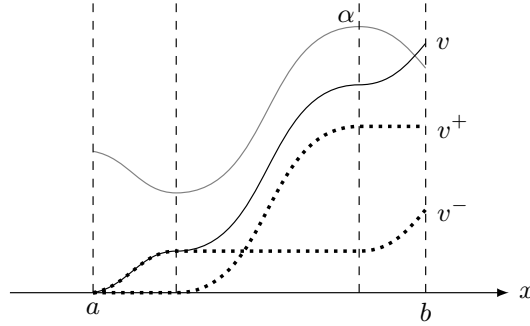
$$v(x) = V_{[a,x]}(\alpha).$$

Intuitively, the change  $\Delta\alpha$  of  $\alpha$  is positive when  $\alpha$  is increasing and negative when  $\alpha$  is decreasing. The variation function  $v(x)$  keeps track of both as positive changes. We wish to introduce the function  $v^+(x)$  that only keeps track of the positive changes (which means constant on the intervals on which  $\alpha$  is decreasing), and also introduce the similar function  $v^-(x)$  that only keeps track of the negative changes. In other words, the two functions should satisfy

$$v^+(x) + v^-(x) = v(x), \quad v^+(x) - v^-(x) = \alpha(x) - \alpha(a).$$

Therefore we define the *positive variation function* and the *negative variation function* by

$$v^+(x) = \frac{v(x) + \alpha(x) - \alpha(a)}{2}, \quad v^-(x) = \frac{v(x) - \alpha(x) + \alpha(a)}{2}.$$



**Figure 4.6.1.** Variation functions.

**Proposition 4.6.2.** *A function  $\alpha$  has bounded variation if and only if it is the difference of two increasing functions. Moreover, the expression  $\alpha(x) = v^+(x) - v^-(x) + \alpha(a)$  as the difference of positive and negative variation functions is the most efficient in the sense that, if  $\alpha = u^+ - u^- + C$  for increasing functions  $u^+, u^-$  and constant  $C$ , then for any  $x < y$ , we have*

$$v^+(y) - v^+(x) \leq u^+(y) - u^+(x), \quad v^-(y) - v^-(x) \leq u^-(y) - u^-(x).$$

*Proof.* Suppose  $\alpha$  has bounded variation. Then  $\alpha(x) = v^+(x) - v^-(x) + \alpha(a)$ . Moreover, for  $x < y$ , we have

$$\begin{aligned} v^+(y) - v^+(x) &= \frac{1}{2}(V_{[a,y]}(\alpha) - V_{[a,x]}(\alpha) + \alpha(y) - \alpha(x)) \\ &= \frac{1}{2}(V_{[x,y]}(\alpha) + \alpha(y) - \alpha(x)) \\ &\geq \frac{1}{2}(V_{[x,y]}(\alpha) - |\alpha(y) - \alpha(x)|) \geq 0. \end{aligned}$$

The first equality is due to the definition of  $v$  and  $v^+$ . The second equality and the second inequality follow from Proposition 4.6.1. The above proves that  $v^+$  is increasing. By similar reason,  $v^-$  is also increasing.

Conversely, monotone functions have bounded variations and linear combinations of bounded variation functions still have bounded variations. So the difference of two increasing functions has bounded variation.

Let  $\alpha = u^+ - u^- + C$  for increasing functions  $u^+, u^-$  and constant  $C$ . Then we have

$$\begin{aligned} v(y) - v(x) &= V_{[x,y]}(\alpha) \leq V_{[x,y]}(u^+) + V_{[x,y]}(u^-) + V_{[x,y]}(C) \\ &= (u^+(y) - u^+(x)) + (u^-(y) - u^-(x)) + 0. \end{aligned}$$

The inequality follows from Proposition 4.6.1, and the second equality is due to  $u^+$  and  $u^-$  increasing. On the other hand, we have

$$\alpha(y) - \alpha(x) = (u^+(y) - u^+(x)) - (u^-(y) - u^-(x)).$$

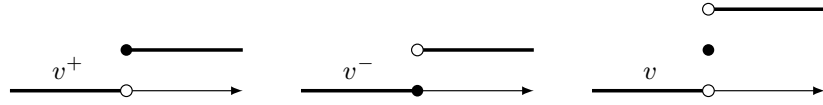
Therefore

$$v^+(y) - v^+(x) = \frac{1}{2}[(v(y) - v(x)) - (\alpha(y) - \alpha(x))] \leq u^+(y) - u^+(x).$$

By similar reason, we also have  $v^-(y) - v^-(x) \leq u^-(y) - u^-(x)$ .  $\square$

**Example 4.6.3.** For the function  $d_c(x)$  in Example 4.1.3, we have

$$v^+(x) = \begin{cases} 0, & \text{if } x \leq c, \\ 1, & \text{if } x > c; \end{cases} \quad v^-(x) = \begin{cases} 0, & \text{if } x < c, \\ 1, & \text{if } x \geq c; \end{cases} \quad v(x) = \begin{cases} 0, & \text{if } x < c, \\ 1, & \text{if } x = c, \\ 2, & \text{if } x > c. \end{cases}$$



**Figure 4.6.2.** Variations of the  $\delta$ -function.

**Exercise 4.73.** Find the positive and negative variation functions.

1. The sign function.
2.  $\sin x$  on  $[0, +\infty)$ .
3.  $f\left(\frac{1}{n}\right) = a_n$  and  $f(x) = 0$  otherwise.

**Exercise 4.74.** Find the positive and negative variation functions of  $\alpha(x) = \int_a^x g(t)dt$ .

**Exercise 4.75.** Suppose  $\alpha = u^+ - u^- + c$  for increasing  $u^+, u^-$  and constant  $c$ . Prove that

$$V_{[a,b]}(\alpha) \leq (u^+(b) - u^+(a)) + (u^-(b) - u^-(a)).$$

Moreover, the equality holds if and only if  $u^+ = v^+ + c^+$  and  $u^- = v^- + c^-$  for some constants  $c^+$  and  $c^-$ .

**Proposition 4.6.3.** Suppose  $v, v^+, v^-$  are variation functions of a bounded variation function  $\alpha$ . Then at any  $c$ , either  $v^+$  or  $v^-$  is right continuous. In particular, the following are equivalent.

1.  $\alpha$  is right continuous at  $c$ .
2.  $v^+$  and  $v^-$  are right continuous at  $c$ .

3.  $v$  is right continuous at  $c$ .

The same equivalence happens to the continuity.

*Proof.* Suppose both  $v^+$  and  $v^-$  are not right continuous at  $c$ . We will construct a more efficient decomposition of  $\alpha$ . Since both  $v^+$  and  $v^-$  are increasing, we have

$$v^+(c^+) - v^+(c) > \epsilon, \quad v^-(c^+) - v^-(c) > \epsilon \text{ for some } \epsilon > 0.$$

Then the functions

$$u^+(x) = \begin{cases} v^+(x), & \text{if } x < c, \\ v^+(x) - \epsilon, & \text{if } x \geq c; \end{cases} \quad u^-(x) = \begin{cases} v^-(x), & \text{if } x < c, \\ v^-(x) - \epsilon, & \text{if } x \geq c, \end{cases}$$

are still increasing and we still have  $\alpha = u^+ - u^- + C$ . However, for  $x < c$ , we have  $u^+(c) - u^+(x) = v^+(c) - \epsilon - v^+(x) < v^+(c) - v^+(x)$ . This violates the most efficiency claim in Proposition 4.6.2.

Thus we conclude that either  $v^+$  or  $v^-$  is right continuous at  $c$ . By  $\alpha = v^+ - v^- + \alpha(a)$  and  $v = v^+ + v^-$ , we further get the equivalence between the three statements.

The similar proof applies to the left continuity, and we get the equivalence for the continuity by combining left and right continuity.  $\square$

**Proposition 4.6.4.** *Suppose  $\alpha$  is a continuous bounded variation function. Then for any  $\epsilon > 0$ , that there is  $\delta > 0$ , such that  $\|P\| < \delta$  implies  $V_P(\alpha) > V_{[a,b]}(\alpha) - \epsilon$ .*

*Proof.* By the definition of variation, for any  $\epsilon_1 > 0$ , there is a partition  $Q$ , such that

$$V_Q(\alpha) > V_{[a,b]}(\alpha) - \epsilon_1.$$

Let  $q$  be the number of partition points in  $Q$ .

For any  $\epsilon_2 > 0$ , by the uniform (see Theorem 2.4.1) continuity of  $\alpha$ , there is  $\delta > 0$ , such that

$$|x - y| < \delta \implies |\alpha(x) - \alpha(y)| < \epsilon_2.$$

Then for any partition  $P$  satisfying  $\|P\| < \delta$ , we have  $|\Delta\alpha_i| < \epsilon_2$  on the intervals in  $P$ . By  $\|P \cup Q\| \leq \|P\| < \delta$ , we also have  $|\Delta\alpha_i| < \epsilon_2$  on the intervals in  $P \cup Q$ . Since  $P \cup Q$  is obtained by adding at most  $q$  points to  $P$ , the  $|\Delta\alpha_i|$  terms in  $V_P(\alpha)$  and in  $V_{P \cup Q}(\alpha)$  are mostly the same, except at most  $q$ -terms in  $V_P(\alpha)$  and at most  $q$ -terms in  $V_{P \cup Q}(\alpha)$ . Therefore we have

$$|V_{P \cup Q}(\alpha) - V_P(\alpha)| \leq 2q\epsilon_2.$$

On the other hand, for the refinement  $P \cup Q$  of  $Q$ , we have

$$V_Q(\alpha) \leq V_{P \cup Q}(\alpha).$$

Therefore

$$V_P(\alpha) > V_{P \cup Q}(\alpha) - 2q\epsilon_2 \geq V_Q(\alpha) - 2q\epsilon_2 > V_{[a,b]}(\alpha) - \epsilon_1 - 2q\epsilon_2.$$



So for any  $\epsilon > 0$ , we start by choosing  $\epsilon = \frac{\epsilon}{2}$ . Then we find a partition  $Q$  and get the number  $q$  of partition points in  $Q$ . Then we choose  $\epsilon_2 = \frac{\epsilon}{4q}$  and find  $\delta$ . By such choices, we get  $\|P\| < \delta$  implying  $V_P(\alpha) > V_{[a,b]}(\alpha) - \epsilon$ .  $\square$

### Criterion for Riemann-Stieltjes Integrability

**Theorem 4.6.5.** *Suppose  $f$  is bounded and  $\alpha$  has bounded variation.*

1. *If  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ , then for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that*

$$\|P\| < \delta \implies \sum \omega_{[x_{i-1}, x_i]}(f) |\Delta \alpha_i| < \epsilon.$$

2. *If for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that*

$$\|P\| < \delta \implies \sum \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha) < \epsilon,$$

*then  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ .*

*Moreover, if  $\alpha$  is monotone or continuous, then both become necessary and sufficient conditions for the Riemann-Stieltjes integrability.*

*Proof.* The proof is very similar to the proof of Theorem 4.1.3. For the first part, we note that by taking  $P = P'$  but choosing different  $x_i^*$  for  $P$  and  $P'$ , we have

$$S(P, f, \alpha) - S(P', f, \alpha) = \sum (f(x_i^*) - f(x_i'^*)) \Delta \alpha_i.$$

Then for each  $i$ , we may choose  $f(x_i^*) - f(x_i'^*)$  to have the same sign as  $\Delta \alpha_i$  and  $|f(x_i^*) - f(x_i'^*)|$  to be as close to the oscillation  $\omega_{[x_{i-1}, x_i]}(f)$  as possible. The result is that  $S(P, f, \alpha) - S(P', f, \alpha)$  is very close to  $\sum \omega_{[x_{i-1}, x_i]}(f) |\Delta \alpha_i|$ . The rest of the proof is the same.

For the second part, the key is that, for a refinement  $Q$  of  $P$ , we have

$$\begin{aligned} |S(P, f, \alpha) - S(Q, f, \alpha)| &= \left| \sum_{i=1}^n f(x_i^*) \Delta \alpha_i - \sum_{i=1}^n S(Q_{[x_{i-1}, x_i]}, f, \alpha) \right| \\ &\leq \sum_{i=1}^n |f(x_i^*) (\alpha(x_{i+1}) - \alpha(x_i)) - S(Q_{[x_{i-1}, x_i]}, f, \alpha)| \\ &\leq \sum_{i=1}^n \omega_{[x_{i-1}, x_i]}(f) V_{Q_{[x_{i-1}, x_i]}}(\alpha) \\ &\leq \sum_{i=1}^n \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha), \end{aligned}$$

where the second inequality follows from (4.6.1). The rest of the proof is the same.

If  $\alpha$  is monotone, then  $|\Delta \alpha_i| = V_{[x_{i-1}, x_i]}(\alpha)$ , so that the two parts are inverse to each other. If  $\alpha$  is continuous, then we prove that the (weaker) criterion in the

first part implies the (stronger) criterion in the second part. This then implies that both parts give necessary and sufficient conditions for the Riemann-Stieltjes integrability.

Since  $\alpha$  is continuous, by Proposition 4.6.4, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$  implies  $V_P(\alpha) > V_{[a,b]}(\alpha) - \epsilon$ . By the criterion in the first part, there is  $\delta' > 0$ , such that  $\|P\| < \delta'$  implies  $\sum \omega_{[x_{i-1}, x_i]}(f) |\Delta \alpha_i| < \epsilon$ . If  $|f| < B$ , then  $\|P\| < \min\{\delta, \delta'\}$  implies

$$\begin{aligned} & \sum \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha) \\ & \leq \sum \omega_{[x_{i-1}, x_i]}(f) |\Delta \alpha_i| + \sum \omega_{[x_{i-1}, x_i]}(f) (V_{[x_{i-1}, x_i]}(\alpha) - |\Delta \alpha_i|) \\ & \leq \epsilon + 2B \sum (V_{[x_{i-1}, x_i]}(\alpha) - |\Delta \alpha_i|) \\ & = \epsilon + 2B(V_{[a,b]}(\alpha) - V_P(\alpha)) < (2B + 1)\epsilon. \end{aligned}$$

This verifies the criterion in the second part. □

**Proposition 4.6.6.** *Any continuous function is Riemann-Stieltjes integrable with respect to any bounded variation function.*

**Proposition 4.6.7.** *Any bounded variation function is Riemann-Stieltjes integrable with respect to any continuous function.*

**Proposition 4.6.8.** *Suppose  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ , which is either monotone or continuous with bounded variation. Suppose  $\phi(y)$  is bounded and uniformly continuous on the values  $f([a, b])$  of  $f(x)$ . Then the composition  $\phi \circ f$  is also Riemann-Stieltjes integrable with respect to  $\alpha$ .*

*Proof.* The proof of Proposition 4.1.4 can be adopted directly to Proposition 4.6.6. Suppose  $f$  is continuous and  $\alpha$  has bounded variation. Then for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $|x - y| < \delta$  implies  $|f(x) - f(y)| < \epsilon$ . This means that for  $\|P\| < \delta$ , we have  $\omega_{[x_{i-1}, x_i]}(f) \leq \epsilon$ . Therefore

$$\sum \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha) \leq \epsilon \sum V_{[x_{i-1}, x_i]}(\alpha) = \epsilon V_{[a,b]}(\alpha).$$

Since  $V_{[a,b]}(\alpha)$  is finite, by the second part of Theorem 4.6.5,  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ .

In the set up of Proposition 4.6.7, we have a bounded variation function  $f$  and a continuous function  $\alpha$ . By Propositions 4.6.6,  $\alpha$  is Riemann-Stieltjes integrable with respect to  $f$ . Then by Proposition 4.5.3,  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ . We cannot use Theorem 4.6.5 to give a direct proof because  $\alpha$  is not assumed to have bounded variation.

Now we prove Proposition 4.6.8 by adopting the proof of Proposition 4.1.6. Note that since  $\alpha$  is either monotone or continuous with bounded variation, Theorem 4.6.5 provides necessary and sufficient condition for the Riemann-Stieltjes integrability.

By the uniform continuity of  $\phi(y)$  on the values of  $f(x)$ , for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\omega_{[c,d]}(f) < \delta$  implies  $\omega_{[c,d]}(\phi \circ f) \leq \epsilon$ . By the Riemann-Stieltjes integrability of  $f$  with respect to  $\alpha$ , for  $\delta\epsilon > 0$ , there is  $\delta' > 0$ , such that

$$\|P\| < \delta' \implies \sum \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha) < \delta\epsilon.$$

Then for those intervals satisfying  $\omega_{[x_{i-1}, x_i]}(f) \geq \delta$ , we have

$$\delta \sum_{\omega_{[x_{i-1}, x_i]}(f) \geq \delta} V_{[x_{i-1}, x_i]}(\alpha) \leq \sum \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha) < \delta\epsilon.$$

If  $\phi$  is bounded by  $B$ , then  $\omega_{[x_{i-1}, x_i]}(\phi \circ f) \leq 2B$ , and we have

$$\sum_{\omega_{[x_{i-1}, x_i]}(f) \geq \delta} \omega_{[x_{i-1}, x_i]}(\phi \circ f) V_{[x_{i-1}, x_i]}(\alpha) \leq 2B \sum_{\omega_{[x_{i-1}, x_i]}(f) \geq \delta} V_{[x_{i-1}, x_i]}(\alpha) < 2B\epsilon.$$

On the other hand, for those intervals with  $\omega_{[x_{i-1}, x_i]}(f) < \delta$ , we have  $\omega_{[x_{i-1}, x_i]}(\phi \circ f) \leq \epsilon$ , so that

$$\sum_{\omega_{[x_{i-1}, x_i]}(f) < \delta} \omega_{[x_{i-1}, x_i]}(\phi \circ f) V_{[x_{i-1}, x_i]}(\alpha) \leq \epsilon \sum_{\omega_{[x_{i-1}, x_i]}(f) < \delta} V_{[x_{i-1}, x_i]}(\alpha) \leq \epsilon V_{[a,b]}(\alpha).$$

Combining the two estimations together, we get

$$\|P\| < \delta' \implies \sum \omega_{[x_{i-1}, x_i]}(\phi \circ f) V_{[x_{i-1}, x_i]}(\alpha) \leq (2B + V_{[a,b]}(\alpha))\epsilon.$$

This implies that  $\phi \circ f$  is Riemann-Stieltjes integrable with respect to  $\alpha$ .  $\square$

**Exercise 4.76.** Suppose  $F(x) = \int_a^x f d\alpha$ ,  $f$  is bounded, and  $\alpha$  has bounded variation. Prove that if  $\alpha$  is continuous at  $x_0$  or  $f(x_0) = 0$ , then  $F$  is continuous at  $x_0$ . This is the converse of Exercise 4.72.

**Exercise 4.77.** Suppose  $\alpha$  is monotone. Prove that the product of functions that are Riemann-Stieltjes integrable with respect to  $\alpha$  is still Riemann-Stieltjes integrable with respect to  $\alpha$ .

**Exercise 4.78.** Suppose  $f$  and  $g$  are bounded. Suppose  $\beta$  is either monotone, or is continuous with bounded variation. Suppose  $g$  is Riemann-Stieltjes integrable with respect to  $\beta$ , and  $\alpha(x) = \int_a^x g d\beta$ . Prove that  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  if and only if  $fg$  is Riemann integrable with respect to  $\beta$ . Moreover, we have

$$\int_a^b f d\alpha = \int_a^b fg d\beta.$$

This extends Theorem 4.5.2.

**Exercise 4.79.** Suppose  $\alpha$  has bounded variation and  $f$  is Riemann-Stieltjes integrable with respect to the variation function  $v$  of  $\alpha$ . Prove that  $f$  is Riemann-Stieltjes integrable with respect to  $\alpha$  and  $|f|$  is Riemann-Stieltjes integrable with respect to  $v$ . Moreover, prove that  $\left| \int_a^b f d\alpha \right| \leq \int_a^b |f| dv$ .

## 4.7 Additional Exercise

### Modified Riemann Sum and Riemann Product

Exercise 4.80. Let  $\phi(t)$  be a function defined near 0. For any partition  $P$  of  $[a, b]$  and choice of  $x_i^*$ , define the “modified Riemann sum”

$$S_\phi(P, f) = \sum_{i=1}^n \phi(f(x_i^*)\Delta x_i).$$

Prove that if  $\phi$  is differentiable at 0 and satisfies  $\phi(0) = 0$ ,  $\phi'(0) = 1$ , then for any integrable  $f(x)$ , we have  $\lim_{\|P\| \rightarrow 0} S_\phi(P, f) = \int_a^b f(x)dx$ .

Exercise 4.81. For any partition  $P$  of  $[a, b]$  and choice of  $x_i^*$ , define the “Riemann product”

$$\Pi(P, f) = (1 + f(x_1^*)\Delta x_1)(1 + f(x_2^*)\Delta x_2) \cdots (1 + f(x_n^*)\Delta x_n).$$

Prove that if  $f(x)$  is integrable, then  $\lim_{\|P\| \rightarrow 0} \Pi(P, f) = e^{\int_a^b f(x)dx}$ .

### Integrability and Continuity

Riemann integrable functions are not necessarily continuous. How much discontinuity can a Riemann integrable function have?

Exercise 4.82. Suppose  $f(x)$  is integrable on  $[a, b]$ . Prove that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that for any partition  $P$  satisfying  $\|P\| < \delta$ , we have  $\omega_{[x_{i-1}, x_i]}(f) < \epsilon$  for some interval  $[x_{i-1}, x_i]$  in the partition.

Exercise 4.83. Suppose there is a sequence of intervals  $[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \cdots$ , such that  $a_n < c < b_n$  for all  $n$  and  $\lim_{n \rightarrow \infty} \omega_{[a_n, b_n]}(f) = 0$ . Prove that  $f(x)$  is continuous at  $c$ .

Exercise 4.84. Prove that an integrable function must be continuous somewhere. In fact, prove that for any  $(c, d) \subset [a, b]$ , an integrable function on  $[a, b]$  is continuous somewhere in  $(c, d)$ . In other words, the continuous points of an integrable function must be dense.

Exercise 4.85. Define the *oscillation* of a function at a point (see Exercises 2.71 through 2.74 for the upper and lower limits of a function)

$$\omega_x(f) = \lim_{\delta \rightarrow 0^+} \omega_{[x-\delta, x+\delta]}(f) = \overline{\lim}_{y \rightarrow x} f(x) - \underline{\lim}_{y \rightarrow x} f(x).$$

Prove that  $f(x)$  is continuous at  $x_0$  if and only if  $\omega_{x_0}(f) = 0$ .

Exercise 4.86. Suppose  $f(x)$  is integrable. Prove that for any  $\epsilon > 0$  and  $\delta > 0$ , the set of those  $x$  with  $\omega_x(f) \geq \delta$  is contained in a union of finitely many intervals, such that the total length of the intervals is  $< \epsilon$ .

Exercise 4.87 (Hankel<sup>24</sup>). Suppose  $f(x)$  is integrable. Prove that for any  $\epsilon > 0$ , the subset of discontinuous points of  $f(x)$  (i.e., those  $x$  with  $\omega_x(f) > 0$ ) is contained in a union of

<sup>24</sup>Hermann Hankel, born 1839 in Halle (Germany), died 1873 in Schramberg (Germany). Hankel was Riemann's student, and his study of Riemann's integral prepared for the discovery of Lebesgue integral.

countably many intervals, such that the total length of the intervals is  $< \epsilon$ . This basically says that the set of discontinuous points of a Riemann integrable function has Lebesgue measure 0. Theorem 10.4.5 says that the converse is also true.

### Strict Inequality in Integration

The existence of continuous points for integrable functions (see Exercise 4.84) enables us to change the inequalities in Proposition 4.3.2 to strict inequalities.

**Exercise 4.88.** Prove that if  $f(x) > 0$  is integrable on  $[a, b]$ , then  $\int_a^b f(x)dx > 0$ . In particular, this shows

$$f(x) < g(x) \implies \int_a^b f(x)dx < \int_a^b g(x)dx.$$

**Exercise 4.89.** Suppose  $f(x)$  is integrable on  $[a, b]$ . Prove that the following are equivalent.

1.  $\int_c^d f(x)dx = 0$  for any  $[c, d] \subset [a, b]$ .
2.  $\int_a^b |f(x)|dx = 0$ .
3.  $\int_a^b f(x)g(x)dx = 0$  for any continuous function  $g(x)$ .
4.  $\int_a^b f(x)g(x)dx = 0$  for any integrable function  $g(x)$ .
5.  $f(x) = 0$  at continuous points.

### Integral Continuity

Exercise 4.24 says that the continuity implies the integral continuity. In fact, the integrability already implies the integral continuity.

**Exercise 4.90.** Suppose  $f$  is integrable on an open interval containing  $[a, b]$ . Suppose  $P$  is a partition of  $[a, b]$  by intervals of equal length  $\delta$ . Prove that if  $|t| < \delta$ , then

$$\int_{x_{i-1}}^{x_i} |f(x+t) - f(x)|dx \leq \delta [\omega_{[x_{i-2}, x_{i-1}]}(f) + \omega_{[x_{i-1}, x_i]}(f) + \omega_{[x_i, x_{i+1}]}(f)].$$

**Exercise 4.91.** Suppose  $f$  is integrable on an open interval containing  $[a, b]$ . Prove that  $\lim_{t \rightarrow 0} \int_a^b |f(x+t) - f(x)|dx = 0$  and  $\lim_{t \rightarrow 1} \int_a^b |f(tx) - f(x)|dx = 0$ .

### Some Integral Inequalities

**Exercise 4.92.** Suppose  $f(x)$  is continuous on  $[a, b]$ , differentiable on  $(a, b)$ , and satisfies  $f(a) = 0$ ,  $0 \leq f'(x) \leq 1$ . Prove that  $\left(\int_a^b f(x)dx\right)^2 \geq \int_a^b f(x)^3 dx$ .

**Exercise 4.93.** Suppose  $f(x)$  has integrable derivative on  $[a, b]$ , and  $f(x) = 0$  somewhere on  $[a, b]$ . Prove that  $|f(x)| \leq \int_a^b |f'(x)| dx$ .

**Exercise 4.94.** Suppose  $f(x)$  has integrable derivative on  $[a, b]$ . Prove that  $\int_a^b |f(x)| dx \leq \max \left\{ \left| \int_a^b f(x) dx \right|, (b-a) \int_a^b |f'(x)| dx \right\}$ .

**Exercise 4.95.** Suppose  $p, q > 0$  satisfy  $\frac{1}{p} + \frac{1}{q} = 1$ . Prove the integral versions of Hölder's and Minkowski's inequalities (see Exercises 3.75 and 3.76).

$$\begin{aligned} \int_a^b |f(x)g(x)| dx &\leq \left( \int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} \left( \int_a^b |g(x)|^q dx \right)^{\frac{1}{q}}, \\ \left( \int_a^b |f(x) + g(x)|^p dx \right)^{\frac{1}{p}} &\leq \left( \int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} + \left( \int_a^b |g(x)|^p dx \right)^{\frac{1}{p}}. \end{aligned}$$

**Exercise 4.96.** Suppose  $f(x)$  has integrable derivative on  $[a, b]$ . Use Hölder's inequality in Exercise 4.95 to prove that

$$(f(b) - f(a))^2 \leq (b-a) \int_a^b f'(x)^2 dx.$$

Then prove the following.

1. If  $f(a) = 0$ , then  $\int_a^b f(x)^2 dx \leq \frac{(b-a)^2}{2} \int_a^b f'(x)^2 dx$ .
2. If  $f(a) = f(b) = 0$ , then  $\int_a^b f(x)^2 dx \leq \frac{(b-a)^2}{4} \int_a^b f'(x)^2 dx$ .
3. If  $f\left(\frac{a+b}{2}\right) = 0$ , then  $\int_a^b f(x)^2 dx \leq \frac{(b-a)^2}{4} \int_a^b f'(x)^2 dx$ .

### Integral Jensen's Inequality

**Exercise 4.97.** Suppose  $f(x)$  is a convex function on  $[a, b]$ . By comparing  $f(x)$  with suitable linear functions, prove that

$$f\left(\frac{a+b}{2}\right)(b-a) \leq \int_a^b f(x) dx \leq \frac{f(a) + f(b)}{2}(b-a).$$

**Exercise 4.98.** A *weight* on  $[a, b]$  is a function  $\lambda(x)$  satisfying

$$\lambda(x) \geq 0, \quad \frac{1}{b-a} \int_a^b \lambda(x) dx = 1.$$

We have  $\frac{1}{b-a} \int_a^b \lambda(x) x dx = (1-\mu)a + \mu b$  for some  $0 < \mu < 1$ . For a convex function on  $[a, b]$ , prove that

$$f((1-\mu)a + \mu b) \leq \frac{1}{b-a} \int_a^b \lambda(x) f(x) dx \leq (1-\mu)f(a) + \mu f(b).$$

The left inequality is the *integral version* of Jensen's inequality in Exercise 3.118. What do you get by applying the integral Jensen's inequality to  $x^2$ ,  $e^x$  and  $\log x$ ?

**Exercise 4.99.** Suppose  $f(x)$  is a convex function on  $[a, b]$  and  $\phi(t)$  is an integrable function on  $[\alpha, \beta]$  satisfying  $a \leq \phi(t) \leq b$ . Suppose  $\lambda(x)$  is a weight function on  $[\alpha, \beta]$  as defined in Exercise 4.98. Prove that

$$f\left(\frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \lambda(t) \phi(t) dt\right) \leq \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \lambda(t) f(\phi(t)) dt.$$

This further extends the integral Jensen's inequality.

**Exercise 4.100.** A special case of Exercise 4.99 is

$$f\left(\int_0^1 \phi(t) dt\right) \leq \int_0^1 f(\phi(t)) dt$$

for any integrable function  $\phi(t)$  on  $[0, 1]$ . Prove the converse that, if the inequality above holds for any  $\phi(t)$  satisfying  $a \leq \phi(t) \leq b$ , then  $f$  is convex on  $[a, b]$ .

### Estimation of Integral

**Exercise 4.101.** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . By comparing  $f(x)$  with the straight line  $L(x) = f(a) + l(x - a)$  for  $l = \sup_{(a,b)} f'$  or  $l = \inf_{(a,b)} f'$ , prove that

$$\frac{\inf_{(a,b)} f'}{2} (b - a)^2 \leq \int_a^b f(x) dx - f(a)(b - a) \leq \frac{\sup_{(a,b)} f'}{2} (b - a)^2.$$

Then use Darboux's Intermediate Value Theorem in Exercise 3.53 to show that

$$\int_a^b f(x) dx = f(a)(b - a) + \frac{f'(c)}{2} (b - a)^2 \text{ for some } c \in (a, b).$$

**Exercise 4.102.** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Suppose  $g(x)$  is non-negative and integrable on  $[a, b]$ . Prove that

$$\int_a^b f(x) g(x) dx = f(a) \int_a^b g(x) dx + f'(c) \int_a^b (x - a) g(x) dx \text{ for some } c \in (a, b).$$

**Exercise 4.103.** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Use Exercise 4.101 to prove that

$$\left| \int_a^b f(x) dx - \frac{f(a) + f(b)}{2} (b - a) \right| \leq \frac{\omega_{(a,b)}(f')}{8} (b - a)^2.$$

In fact, this estimation can also be derived from Exercise 4.21.

**Exercise 4.104.** Suppose  $f(x)$  is continuous on  $[a, b]$  and has second order derivative on  $(a, b)$ . Use the Taylor expansion at  $\frac{a+b}{2}$  to prove that

$$\int_a^b f(x) dx = f\left(\frac{a+b}{2}\right) (b - a) + \frac{f''(c)}{24} (b - a)^3 \text{ for some } c \in (a, b).$$

### High Order Estimation of Integral

**Exercise 4.105.** Suppose  $f(x)$  is continuous on  $[a, b]$  and has  $n$ -th order derivative on  $(a, b)$ . By either integrating the Taylor expansion at  $a$  or considering the Taylor expansion of the function  $F(x) = \int_a^x f(t)dt$ , prove that

$$\int_a^b f(x)dx = f(a)(b-a) + \frac{f'(a)}{2!}(b-a)^2 + \cdots + \frac{f^{(n-1)}(a)}{n!}(b-a)^n + \frac{f^{(n)}(c)}{(n+1)!}(b-a)^{n+1}$$

for some  $c \in (a, b)$ .

**Exercise 4.106.** Suppose  $f(x)$  is continuous on  $[a, b]$  and has  $n$ -th order derivative on  $(a, b)$ . Prove that

$$\left| \int_a^b f(x)dx - \sum_{k=0}^{n-1} \frac{f^{(k)}(a) + (-1)^k f^{(k)}(b)}{(k+1)!2^{k+1}} (b-a)^{k+1} \right| \leq \frac{\omega_{(a,b)}(f^{(n)})}{(n+1)!2^{n+1}} (b-a)^{n+1}$$

for odd  $n$ , and

$$\int_a^b f(x)dx = \sum_{k=0}^{n-1} \frac{f^{(k)}(a) + (-1)^k f^{(k)}(b)}{(k+1)!2^{k+1}} (b-a)^{k+1} + \frac{f^{(n)}(c)}{(n+1)!2^n} (b-a)^{n+1}$$

for even  $n$  and some  $c \in (a, b)$ .

**Exercise 4.107.** Suppose  $f(x)$  is continuous on  $[a, b]$  and has  $2n$ -th order derivative on  $(a, b)$ . Prove that

$$\int_a^b f(x)dx = \sum_{k=0}^{n-1} \frac{1}{(2k+1)!2^{2k}} f^{(2k)}\left(\frac{a+b}{2}\right) (b-a)^{2k+1} + \frac{f^{(2n)}(c)}{(2n+1)!2^{2n}} (b-a)^{2n+1}$$

for some  $c \in (a, b)$ .

### Estimation of Special Riemann Sums

Consider the partition of  $[a, b]$  by evenly distributed partition points  $x_i = a + \frac{i}{n}(b-a)$ . We can form the Riemann sums by choosing the left, right and middle points of the partition intervals.

$$\begin{aligned} S_{\text{left},n}(f) &= \sum f(x_{i-1})\Delta x_i = \frac{b-a}{n} \sum f(x_{i-1}), \\ S_{\text{right},n}(f) &= \sum f(x_i)\Delta x_i = \frac{b-a}{n} \sum f(x_i), \\ S_{\text{middle},n}(f) &= \sum f\left(\frac{x_{i-1} + x_i}{2}\right) \Delta x_i = \frac{b-a}{n} \sum f\left(\frac{x_{i-1} + x_i}{2}\right). \end{aligned}$$

The question is how close these are to the actual integral.



**Exercise 4.108.** Suppose  $f(x)$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ , such that  $f'(x)$  is integrable. Use the estimation in Exercise 4.100 to prove that

$$\begin{aligned}\lim_{n \rightarrow \infty} n \left( \int_a^b f(x) dx - S_{\text{left},n}(f) \right) &= \frac{1}{2}(f(b) - f(a))(b - a), \\ \lim_{n \rightarrow \infty} n \left( \int_a^b f(x) dx - S_{\text{right},n}(f) \right) &= -\frac{1}{2}(f(b) - f(a))(b - a).\end{aligned}$$

**Exercise 4.109.** Suppose  $f(x)$  is continuous on  $[a, b]$  and has second order derivative on  $(a, b)$ , such that  $f''(x)$  is integrable on  $[a, b]$ . Use the estimation in Exercise 4.104 to prove that

$$\lim_{n \rightarrow \infty} n^2 \left( \int_a^b f(x) dx - S_{\text{middle},n}(f) \right) = \frac{1}{24}(f'(b) - f'(a))(b - a)^2.$$

**Exercise 4.110.** Use Exercises 4.105, 4.106 and 4.107 to derive higher order approximation formulae for the integral  $\int_a^b f(x) dx$ .

### Average of Function

The average of an integrable function on  $[a, b]$  is

$$\text{Av}_{[a,b]}(f) = \frac{1}{b-a} \int_a^b f(x) dx.$$

**Exercise 4.111.** Prove the properties of the average.

1.  $\text{Av}_{[a+c, b+c]}(f(x+c)) = \text{Av}_{[a,b]}(f(x))$ ,  $\text{Av}_{[\lambda a, \lambda b]}(f(\lambda x)) = \text{Av}_{[a,b]}(f(x))$ .
2. If  $c = \lambda a + (1-\lambda)b$ , then  $\text{Av}_{[a,b]}(f) = \lambda \text{Av}_{[a,c]}(f) + (1-\lambda) \text{Av}_{[c,b]}(f)$ . In particular,  $\text{Av}_{[a,b]}(f)$  lies between  $\text{Av}_{[a,c]}(f)$  and  $\text{Av}_{[c,b]}(f)$ .
3.  $f \geq g$  implies  $\text{Av}_{[a,b]}(f) \geq \text{Av}_{[a,b]}(g)$ .
4. If  $f(x)$  is continuous, then  $\text{Av}_{[a,b]}(f) = f(c)$  for some  $c \in (a, b)$ .

**Exercise 4.112.** Suppose  $f(x)$  is integrable on  $[0, a]$  for any  $a > 0$ . Consider the average function  $g(x) = \text{Av}_{[0,x]}(f) = \frac{1}{x} \int_0^x f(t) dt$ .

1. Prove that if  $\lim_{x \rightarrow +\infty} f(x) = l$ , then  $\lim_{x \rightarrow +\infty} g(x) = l$  (compare Exercise 1.66).
2. Prove that if  $f(x)$  is increasing, then  $g(x)$  is also increasing.
3. Prove that if  $f(x)$  is convex, then  $g(x)$  is also convex.

**Exercise 4.113.** For a weight function  $\lambda(x)$  in Exercise 4.97, the *weighted average* of an integrable function  $f(x)$  is

$$\text{Av}_{[a,b]}^\lambda(f(x)) = \frac{1}{b-a} \int_a^b \lambda(x) f(x) dx$$

Can you extend the properties of average in Exercise 4.111 to weighted average?

**Second Integral Mean Value Theorem**

Exercise 4.114. Suppose  $f \geq 0$  and is decreasing. Suppose  $m \leq \int_a^x g(x)dx \leq M$  for  $x \in [a, b]$ . Prove that

$$f(a)m \leq \int_a^b f(x)g(x)dx \leq f(a)M.$$

Then use this to prove that

$$\int_a^b f(x)g(x)dx = f(a) \int_a^c g(x)dx \text{ for some } c \in (a, b).$$

What if  $f(x)$  is increasing?

Exercise 4.115. Suppose  $f$  is monotone. Prove that

$$\int_a^b f(x)g(x)dx = f(a) \int_a^c g(x)dx + f(b) \int_c^b g(x)dx \text{ for some } c \in (a, b).$$

Exercise 4.116. Suppose  $f \geq 0$  is decreasing and is Riemann-Stieltjes integrable with respect to  $\alpha$ . Suppose  $m \leq \alpha \leq M$  on  $[a, b]$ . Prove that

$$f(a)(m - \alpha(a)) \leq \int_a^b f d\alpha \leq f(a)(M - \alpha(a)).$$

This extends Exercise 4.114. Can you also extend Exercise 4.115 to the Riemann-Stieltjes integral?

## Chapter 5

# Topics in Analysis

## 5.1 Improper Integration

The Riemann integral was defined only for a bounded function on a bounded interval. An integration becomes *improper* when the function or the interval becomes unbounded. Improper integrals can be evaluated by first integrating on the bounded part and then take limit. For example, if  $f$  is integrable on  $[a, c]$  for any

$a < c < b$ , and  $\lim_{c \rightarrow b^-} \int_a^c f(x) dx$  converges, then we say that the improper integral  $\int_a^b f(x) dx$  converges and write

$$\int_a^b f(x) dx = \lim_{c \rightarrow b^-} \int_a^c f(x) dx.$$

The definition also applies to the case  $b = +\infty$ , and similar definition can be made when  $f(x)$  is integrable on  $[c, b]$  for any  $a < c < b$ .

By Exercise 4.7, if  $f(x)$  is a bounded function on bounded interval  $[a, b]$  and is integrable on  $[a, c]$  for any  $c \in (a, b)$ , then it is integrable on  $[a, b]$ . Therefore an integral becomes improper at  $b$  only if  $b = \infty$  or  $b$  is finite but  $f$  is unbounded near  $b$ .

**Example 5.1.1.** The integral  $\int_0^1 x^p dx$  is improper at  $0^+$  for  $p < 0$ . For  $0 < c < 1$ , we have  $\int_c^1 x^p dx = \frac{1 - c^{p+1}}{p+1}$ . As  $c \rightarrow 0^+$ , this converges if and only if  $p+1 > 0$ . Therefore  $\int_0^1 x^p dx$  converges if and only if  $p > -1$ , and  $\int_0^1 x^p dx = \frac{1}{p+1}$ .

Similarly, the integral  $\int_1^{+\infty} x^p dx$  is improper at  $+\infty$  and converges if and only if  $p < -1$ . Moreover, we have  $\int_1^{+\infty} x^p dx = \frac{-1}{p+1}$  for  $p < -1$ .

The integral  $\int_0^{+\infty} x^p dx$  is improper at  $0^+$  for  $p < 0$  and is always improper at  $+\infty$ . Because the convergence at  $0^+$  (which requires  $p > -1$ ) and the convergence at  $+\infty$  (which requires  $p < -1$ ) are contradictory, the integral  $\int_0^{+\infty} x^p dx$  never converges.

**Example 5.1.2.** The integral  $\int_0^1 \log x dx$  is improper at  $0^+$ . For any  $0 < c < 1$ , we have

$$\int_c^1 \log x dx = 1 \log 1 - c \log c - \int_c^1 x d \log x = -c \log c - \int_c^1 dx = -c \log c - 1 + c.$$

Therefore

$$\int_0^1 \log x dx = \lim_{c \rightarrow 0^+} (-c \log c - 1 + c) = 1$$

converges.

The example shows that the integration by parts can be extended to improper integrals by taking the limit of the integration by parts for proper integrals. By the same reason, the change of variable can also be extended to improper integrals.

**Exercise 5.1.** Suppose  $f(x)$  is continuous for  $x \geq 0$ , and  $\lim_{x \rightarrow +\infty} f(x) = l$ . Prove that for any  $a, b > 0$ , we have  $\int_0^{+\infty} \frac{f(ax) - f(bx)}{x} dx = (f(0) - l) \log \frac{b}{a}$ .

**Exercise 5.2.** Prove that

$$\int_0^{+\infty} f\left(ax + \frac{b}{x}\right) dx = \frac{1}{a} \int_0^{+\infty} f(\sqrt{x^2 + 4ab}) dx,$$

provided  $a, b > 0$  and both sides converge.

**Exercise 5.3.** Formulate general theorem on the integration by parts and change of variable for improper integral.

## Convergence Test

Let  $a$  be fixed and let  $f$  be integrable on  $[a, c]$  for any  $c > a$ . Then  $\int_a^{+\infty} f(x) dx$  is improper at  $+\infty$ . The Cauchy criterion for the convergence of the improper integral is that, for any  $\epsilon > 0$ , there is  $N$ , such that

$$b, c > N \implies \left| \int_a^b f(x) dx - \int_a^c f(x) dx \right| = \left| \int_b^c f(x) dx \right| < \epsilon.$$

Similar Cauchy criterion can be made for other types of improper integrals.

An immediate consequence of the Cauchy criterion is the following test for convergence, again stated only for improper integral at  $+\infty$ .

**Proposition 5.1.1 (Comparison Test).** Suppose  $f(x)$  and  $g(x)$  are integrable on  $[a, b]$  for any  $b > a$ . If  $|f(x)| \leq g(x)$  and  $\int_a^{+\infty} g(x) dx$  converges, then  $\int_a^{+\infty} f(x) dx$  also converges.

*Proof.* By the Cauchy criterion for the convergence of  $\int_a^{+\infty} g(x) dx$ , For any  $\epsilon > 0$ , there is  $N$ , such that  $b, c > N$  implies  $\int_b^c g(x) dx < \epsilon$ . Then

$$b, c > N \implies \left| \int_b^c f(x) dx \right| \leq \int_b^c |f(x)| dx \leq \int_b^c g(x) dx < \epsilon.$$

This verifies the Cauchy criterion for the convergence of  $\int_a^{+\infty} f(x) dx$ .  $\square$

A special case of the comparison test is  $g(x) = |f(x)|$ , which shows that the convergence of  $\int_a^{+\infty} |f(x)| dx$  implies the convergence of  $\int_a^{+\infty} f(x) dx$ . Therefore there are three possibilities for an improper integral.

1. *Absolutely convergent:*  $\int_a^{+\infty} |f(x)|dx$  converges. Therefore  $\int_a^{+\infty} f(x)dx$  also converges.
2. *Conditionally convergent:*  $\int_a^{+\infty} f(x)dx$  converges and  $\int_a^{+\infty} |f(x)|dx$  diverges.
3. *Divergent:*  $\int_a^{+\infty} f(x)dx$  diverges. Therefore  $\int_a^{+\infty} |f(x)|dx$  also diverges.

**Example 5.1.3.** The integral  $\int_1^{+\infty} \frac{\sin x}{x} dx$  is improper at  $+\infty$ . The integral of the corresponding absolute value function satisfies

$$\int_1^{n\pi} \left| \frac{\sin x}{x} \right| dx = \sum_{k=2}^n \int_{(k-1)\pi}^{k\pi} \left| \frac{\sin x}{x} \right| dx \geq \sum_{k=2}^n \frac{1}{k\pi} \int_{(k-1)\pi}^{k\pi} |\sin x| dx = \frac{2}{k\pi} \sum_{k=2}^n \frac{1}{k}.$$

By Example 1.4.4, the right side diverges to  $+\infty$  as  $n \rightarrow +\infty$ . Therefore  $\int_1^a \left| \frac{\sin x}{x} \right| dx$  is not bounded, and  $\int_1^{+\infty} \left| \frac{\sin x}{x} \right| dx$  diverges.

On the other hand, for  $b > 0$ , we use the integration by parts to get

$$\int_1^b \frac{\sin x}{x} dx = - \int_1^b \frac{1}{x} d \cos x = \frac{\sin b}{b} - \frac{\sin 1}{1} + \int_1^b \frac{\cos x}{x^2} dx.$$

This implies

$$\int_1^{+\infty} \frac{\sin x}{x} dx = - \int_1^{+\infty} \frac{1}{x} d \cos x = - \frac{\sin 1}{1} + \int_1^{+\infty} \frac{\cos x}{x^2} dx,$$

in the sense that the left side converges if and only if the right side converges, and both sides have the same value when they converge. By  $\left| \frac{\cos x}{x^2} \right| \leq \frac{1}{x^2}$ , the convergence of  $\int_1^{+\infty} \frac{dx}{x^2}$  (see Example 5.1.1), and the comparison test Theorem 5.1.1, the improper integral  $\int_1^{+\infty} \frac{\cos x}{x^2} dx$  converges. Therefore  $\int_1^{+\infty} \frac{\sin x}{x} dx$  converges.

We conclude that  $\int_1^{+\infty} \frac{\sin x}{x} dx$  conditionally converges.

**Exercise 5.4.** Suppose  $f(x) \leq g(x) \leq h(x)$ . Prove that if  $\int_a^b |f(x)|dx$  and  $\int_a^b |h(x)|dx$  converge, then  $\int_a^b g(x)dx$  converges.

**Exercise 5.5.** Prove that if  $f(x) \geq 0$  and  $\int_a^{+\infty} f(x)dx$  converges, then there is an increasing sequence  $x_n$  diverging to  $+\infty$ , such that  $\lim_{n \rightarrow \infty} f(x_n) = 0$ . Moreover, prove that in the special case  $f(x)$  is monotone, we have  $\lim_{x \rightarrow +\infty} f(x) = 0$ .

**Exercise 5.6.** Prove that for a bounded interval  $[a, b]$ , if  $\int_a^b f(x)^2 dx$  converges, then  $\int_a^b f(x) dx$  also converges.

**Exercise 5.7.** Suppose  $f(x) \geq 0$  and  $\lim_{x \rightarrow +\infty} f(x) = 0$ . Prove that if  $\int_a^{+\infty} f(x) dx$  converges, then  $\int_a^{+\infty} f(x)^2 dx$  also converges.

**Exercise 5.8.** Suppose  $f$  is positive and increasing on  $[0, +\infty)$ . Suppose  $F(x) = \int_0^x f(t) dt$ . Prove that  $\int_0^{+\infty} \frac{dx}{f(x)}$  converges if and only if  $\int_0^{+\infty} \frac{x dx}{F(x)}$  converges.

**Exercise 5.9.** Suppose  $f(x)$  is integrable on  $[0, a]$  and continuous at 0. Suppose  $\int_0^{+\infty} |g(x)| dx$  converges. Prove that

$$\lim_{t \rightarrow +\infty} \int_0^{ta} g(x) f\left(\frac{x}{t}\right) dx = f(0) \int_0^{+\infty} g(x) dx.$$

The idea in Example 5.1.3 can be generalised to the following tests.

**Proposition 5.1.2** (Dirichlet Test). *Suppose  $f(x)$  is monotone and  $\lim_{x \rightarrow +\infty} f(x) = 0$ . Suppose there is  $B$ , such that  $\left| \int_a^b g(x) dx \right| < B$  for all  $b \in [a, +\infty)$ . Then  $\int_a^{+\infty} f(x)g(x) dx$  converges.*

**Proposition 5.1.3** (Abel<sup>25</sup> Test). *Suppose  $f(x)$  is monotone and bounded. Suppose  $\int_a^{+\infty} g(x) dx$  converges. Then  $\int_a^{+\infty} f(x)g(x) dx$  converges.*

*Proof.* Let  $G(x) = \int_a^x g(t) dt$ . Then by Theorems 4.5.2 and 4.5.3, we have

$$\int_a^b f g dx = \int_a^b f dG = f(b)G(b) - f(a)G(a) - \int_a^b G df.$$

Under the assumption of either test, we know  $G$  is bounded by a constant  $B$ ,  $\lim_{b \rightarrow +\infty} f(b)$  converges, and  $\lim_{b \rightarrow +\infty} f(b)G(b)$  converges. Therefore the convergence of  $\lim_{b \rightarrow +\infty} \int_a^b f g dx$  is reduced to the convergence of  $\lim_{b \rightarrow +\infty} \int_a^b G df$ . Since

<sup>25</sup>Niels Henrik Abel, born 1802 in Frindoe (Norway), died 1829 in Froland (Norway). In 1824, Abel proved the impossibility of solving the general equation of fifth degree in radicals. Abel also made contributions to elliptic functions. Abel's name is enshrined in the term *abelian*, which describes the commutative property.

$f$  is monotone, by Exercise 4.61, we have

$$\left| \int_b^c G df \right| \leq B |f(b) - f(c)|.$$

Then the Cauchy criterion for the convergence of  $\lim_{b \rightarrow +\infty} f(b)$  implies the Cauchy criterion for the convergence of  $\lim_{b \rightarrow +\infty} \int_a^b G df$ .  $\square$

**Exercise 5.10.** Prove that  $\lim_{a \rightarrow +\infty} a^q \int_a^{+\infty} \frac{\sin x dx}{x^p} = 0$  when  $p > q > 0$ . Then prove that  $\lim_{a \rightarrow 0} \frac{1}{a} \int_0^a \sin \frac{1}{x} dx = 0$ .

**Exercise 5.11.** Derive the Abel test from the Dirichlet test.

**Exercise 5.12.** State the Dirichlet and Abel tests for other kinds of improper integrals, such as unbounded function on bounded interval.

## 5.2 Series of Numbers

A *series* (of numbers) is an infinite sum

$$\sum_{n=1}^{\infty} x_n = x_1 + x_2 + \cdots + x_n + \cdots.$$

The *partial sum* of the series is the sequence

$$s_n = \sum_{k=1}^n x_k = x_1 + x_2 + \cdots + x_n.$$

We say the series converges to *sum*  $s$ , and denote  $\sum_{n=1}^{\infty} x_n = s$ , if  $\lim_{n \rightarrow \infty} s_n = s$ . If  $\lim_{n \rightarrow \infty} s_n = \infty$ , then the series *diverges to infinity*, and  $\sum_{n=1}^{\infty} x_n = \infty$ .

Like sequences, a series does not have to start with index 1. On the other hand, we may always assume that a series starts with index 1 in theoretical studies. Moreover, modifying, adding or deleting finitely many terms in a series does not change the convergence, but may change the sum.

**Example 5.2.1.** The *geometric series* is  $\sum_{n=0}^{\infty} a^n = 1 + a + a^2 + \cdots + a^n + \cdots$  has partial sum  $s_n = \frac{1 - a^{n+1}}{1 - a}$ , and

$$\sum_{n=0}^{\infty} a^n = \begin{cases} \frac{1}{1-a}, & \text{if } |a| < 1, \\ \text{diverges,} & \text{if } |a| \geq 1. \end{cases}$$



**Example 5.2.2.** By Example 1.4.4, we know the *harmonic series*  $\sum_{n=1}^{\infty} \frac{1}{n}$  diverges to  $+\infty$ , and the series  $\sum_{n=1}^{\infty} \frac{1}{n^2}$  and  $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$  converge.

In general, a *non-negative series* has increasing partial sum, and the series converges if and only if the partial sum is bounded.

**Example 5.2.3.** In Example 3.4.9, the estimation of the remainder of the Taylor series tells us that  $\sum_{n=0}^{\infty} \frac{x^n}{n!}$  converges to  $e^x$  for any  $x$ . In particular, we have

$$1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots = e.$$

**Exercise 5.13.** Prove that  $x_n$  converges if and only if  $\sum(x_{n+1} - x_n)$  converges.

**Exercise 5.14.** Prove that if  $\sum x_n$  and  $\sum y_n$  converge, then  $\sum(ax_n + by_n)$  converges, and  $\sum(ax_n + by_n) = a \sum x_n + b \sum y_n$ .

**Exercise 5.15.** For strictly increasing  $n_k$ , the series  $\sum_{k=1}^{\infty} (x_{n_k} + x_{n_k+1} + \cdots + x_{n_{k+1}-1})$  is obtained from  $\sum_{n=1}^{\infty} x_n$  by combining successive terms.

1. Prove that if  $\sum_{n=1}^{\infty} x_n$  converges, then  $\sum_{k=1}^{\infty} (x_{n_k} + x_{n_k+1} + \cdots + x_{n_{k+1}-1})$  also converges.
2. Prove that if the terms  $x_{n_k}, x_{n_k+1}, \dots, x_{n_{k+1}-1}$  that got combined have the same sign and  $\sum_{k=1}^{\infty} (x_{n_k} + x_{n_k+1} + \cdots + x_{n_{k+1}-1})$  converges, then  $\sum_{n=1}^{\infty} x_n$  also converges.

**Exercise 5.16.** Prove that if  $\lim_{n \rightarrow \infty} x_n = 0$ , then  $\sum x_n$  converges if and only if  $\sum(x_{2n-1} + x_{2n})$  converges, and the two sums are the same. What about combining three consecutive terms?

**Exercise 5.17.** Suppose  $x_n$  is decreasing and positive. Prove that  $\sum x_n$  converges if and only if  $\sum 2^n x_{2^n}$  converges. Then study the convergence of  $\sum \frac{1}{n^p}$  and  $\sum \frac{1}{n(\log n)^p}$ .

**Exercise 5.18.** Prove that if  $y$  is not an integer multiple of  $2\pi$ , then

$$\begin{aligned} \sum_{k=0}^n \cos(x + ky) &= \frac{1}{2 \sin \frac{y}{2}} \left[ \sin \left( x + \frac{2n+1}{2}y \right) - \sin \left( x - \frac{1}{2}y \right) \right], \\ \sum_{k=0}^n \sin(x + ky) &= \frac{1}{2 \sin \frac{y}{2}} \left[ -\cos \left( x + \frac{2n+1}{2}y \right) + \cos \left( x - \frac{1}{2}y \right) \right]. \end{aligned}$$

Use integration on  $[\pi, y]$  to find the partial sum of the series  $\sum_{n=1}^{\infty} \frac{\sin ny}{n}$  for  $y \in [0, 2\pi]$ . Then apply Riemann-Lebesgue Lemma in Exercise 4.114 to get  $\sum_{n=1}^{\infty} \frac{\sin ny}{n} = \frac{\pi - y}{2}$  for  $y \in [0, 2\pi]$ .

## Comparison Test

The Cauchy criterion for the convergence of a series  $\sum x_n$  is that, for any  $\epsilon > 0$ , there is  $N$ , such that

$$n \geq m > N \implies |s_n - s_{m-1}| = |x_m + x_{m+1} + \cdots + x_n| < \epsilon.$$

The special case  $m = n$  means that the convergence of  $\sum x_n$  implies  $\lim_{n \rightarrow \infty} x_n = 0$ . Similar to improper integral, the Cauchy criterion leads to the comparison test.

**Proposition 5.2.1** (Comparison Test). *Suppose  $|x_n| \leq y_n$  for sufficiently big  $n$ . If  $\sum y_n$  converges, then  $\sum x_n$  converges.*

Again similar to improper integral, the comparison test leads further to three possibilities for a series.

1. *Absolutely convergent:*  $\sum |x_n|$  converges. Therefore  $\sum x_n$  also converges.
2. *Conditionally convergent:*  $\sum x_n$  converges and  $\sum |x_n|$  diverges.
3. *Divergent:*  $\sum x_n$  diverges. Therefore  $\sum |x_n|$  also diverges.

Exercise 5.19. Prove the comparison test Theorem 5.2.1.

Exercise 5.20. Is there any relation between the convergence of  $\sum x_n$ ,  $\sum y_n$ ,  $\sum \max\{x_n, y_n\}$  and  $\sum \min\{x_n, y_n\}$ ?

Exercise 5.21. Suppose  $x_n > 0$  and  $x_n$  is increasing. Prove that  $\sum \frac{1}{x_n}$  converges if and only if  $\sum \frac{n}{x_1 + x_2 + \cdots + x_n}$  converges.

Exercise 5.22. Suppose  $x_n > 0$ . Prove that  $\sum x_n$  converges if and only if  $\sum \frac{x_n}{x_1 + x_2 + \cdots + x_n}$  converges.

Exercise 5.23. Suppose  $x_n \geq 0$ . Prove that  $\sum \frac{x_n}{(x_1 + x_2 + \cdots + x_n)^2}$  converges.

Exercise 5.24. Suppose  $x_n$  decreases and converges to 0. Prove that if  $\sum_{i=1}^n (x_i - x_n) = \sum_{i=1}^n x_i - nx_n \leq B$  for a fixed bound  $B$ , then  $\sum x_n$  converges.

Exercise 5.25 (Root Test). By comparing with the geometric series in Example 5.2.1, derive the *root test*:

1. If  $\sqrt[n]{|x_n|} \leq a$  for some constant  $a < 1$  and sufficiently big  $n$ , then  $\sum x_n$  converges.
2. If  $\sqrt[n]{|x_n|} \geq 1$  for infinitely many  $n$ , then  $\sum x_n$  diverges.

Then further derive the limit version of the root test:

1. If  $\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|x_n|} < 1$ , then  $\sum x_n$  converges.
2. If  $\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|x_n|} > 1$ , then  $\sum x_n$  diverges.

**Exercise 5.26 (Ratio Test).** Suppose  $\left| \frac{x_{n+1}}{x_n} \right| \leq \frac{y_{n+1}}{y_n}$  for sufficiently big  $n$ . Prove that the convergence of  $\sum y_n$  implies the convergence of  $\sum x_n$ .

**Exercise 5.27.** By taking  $y_n = a^n$  in Exercise 5.26, derive the usual version of the *ratio test*:

1. If  $\left| \frac{x_{n+1}}{x_n} \right| \leq a$  for some constant  $a < 1$  and sufficiently big  $n$ , then  $\sum x_n$  converges.
2. If  $\left| \frac{x_{n+1}}{x_n} \right| \geq 1$  for sufficiently big  $n$ , then  $\sum x_n$  diverges.

Then further derive the limit version of the root test:

1. If  $\overline{\lim}_{n \rightarrow \infty} \left| \frac{x_{n+1}}{x_n} \right| < 1$ , then  $\sum x_n$  converges.
2. If  $\underline{\lim}_{n \rightarrow \infty} \left| \frac{x_{n+1}}{x_n} \right| > 1$ , then  $\sum x_n$  diverges.

Moreover, explain that the lower limit in the last statement cannot be changed to upper limit.

**Exercise 5.28.** Prove that

$$\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|x_n|} \leq \overline{\lim}_{n \rightarrow \infty} \left| \frac{x_{n+1}}{x_n} \right|, \quad \underline{\lim}_{n \rightarrow \infty} \sqrt[n]{|x_n|} \geq \underline{\lim}_{n \rightarrow \infty} \left| \frac{x_{n+1}}{x_n} \right|.$$

What do the inequalities tell you about the relation between the root and ratio tests?

**Exercise 5.29 (Raabe<sup>26</sup> Test).** By taking  $y_n = \frac{1}{(n-1)^p}$  in Exercise 5.26 and use the fact the  $\sum y_n$  converges if and only if  $p > 1$  (to be established after Proposition 5.2.2), derive the *Raabe test*:

1. If  $\left| \frac{x_{n+1}}{x_n} \right| \leq 1 - \frac{p}{n}$  for some constant  $p > 1$  and sufficiently big  $n$ , then  $\sum x_n$  converges.
2. If  $\left| \frac{x_{n+1}}{x_n} \right| \geq 1 - \frac{1}{n}$  for sufficiently big  $n$ , then  $\sum x_n$  diverges. (In fact,  $\left| \frac{x_{n+1}}{x_n} \right| \geq 1 - \frac{1}{n-a}$  is enough.)

Then further derive the limit version of the Raabe test:

1. If  $\underline{\lim}_{n \rightarrow \infty} n \left( 1 - \left| \frac{x_{n+1}}{x_n} \right| \right) > 1$ , then  $\sum x_n$  absolutely converges.
2. If  $\overline{\lim}_{n \rightarrow \infty} n \left( 1 - \left| \frac{x_{n+1}}{x_n} \right| \right) < 1$ , then  $\sum |x_n|$  diverges.

What can happen when the limit is 1?

**Exercise 5.30.** Rephrase the Raabe test in Exercise 5.29 in terms of the quotient  $\left| \frac{x_n}{x_{n+1}} \right|$  and find the corresponding limit version.

<sup>26</sup>Joseph Ludwig Raabe, born 1801 in Brody (now Ukraine), died 1859 in Zürich (Switzerland).

**Exercise 5.31.** Show that the number of  $n$  digit numbers that do not contain the digit 9 is  $8 \cdot 9^{n-1}$ . Then use the fact to prove that if we delete the terms in the harmonic series that contain the digit 9, then the series becomes convergent. What about deleting the terms that contain some other digit? What about the numbers expressed in the base other than 10? What about deleting similar terms in the series  $\sum \frac{1}{n^p}$ ?

## Improper Integral and Series

The convergence of improper integral and the convergence of series are quite similar. There are many ways we can exploit such similarity.

**Proposition 5.2.2 (Integral Comparison Test).** *Suppose  $f(x)$  is a decreasing function on  $[a, +\infty)$  satisfying  $\lim_{x \rightarrow +\infty} f(x) = 0$ . Then the series  $\sum f(n)$  converges if and only if the improper integral  $\int_a^{+\infty} f(x)dx$  converges.*

*Proof.* Since the convergence is not changed if finitely many terms are modified or deleted, we may assume  $a = 1$  without loss of generality.

Since  $f(x)$  is decreasing, we have  $f(k) \geq \int_k^{k+1} f(x)dx \geq f(k+1)$ . Then

$$\begin{aligned} f(1) + f(2) + \cdots + f(n-1) &\geq \int_1^n f(x)dx = \int_1^2 f(x)dx + \int_2^3 f(x)dx + \cdots + \int_{n-1}^n f(x)dx \\ &\geq f(2) + f(3) + \cdots + f(n). \end{aligned}$$

This implies that  $\int_1^n f(x)dx$  is bounded if and only if the partial sums of the series  $\sum f(n)$  are bounded. Since  $f(x) \geq 0$ , the boundedness is equivalent to the convergence. Therefore  $\int_a^{+\infty} f(x)dx$  converges if and only if  $\sum f(n)$  converges.  $\square$

For example, the series  $\sum \frac{1}{(n+a)^p}$  converges if and only if  $\int_b^{+\infty} \frac{dx}{(x+a)^p}$  converges, which means  $p > 1$ . The estimation in the proof can be used in many other ways. See the exercises below and Exercises 5.95 through 5.100.

**Exercise 5.32.** In the setting of Proposition 5.2.2, prove that

$$d_n = f(1) + f(2) + \cdots + f(n) - \int_1^n f(x)dx$$

is decreasing and satisfies  $f(n) \leq d_n \leq f(1)$ . This implies that  $d_n$  converges to a number in  $[0, f(1)]$ . A special case is

$$\lim_{n \rightarrow \infty} \left( 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n \right) = 0.577215669015328 \cdots$$

The number is called the *Euler<sup>27</sup>-Mascheroni<sup>28</sup> constant*.

**Exercise 5.33.** Prove  $\log(n-1)! < n(\log n - 1) + 1 < \log n!$  and then derive the inequality  $\frac{n^n}{e^{n-1}} < n! < \frac{(n+1)^{n+1}}{e^n}$ .

**Exercise 5.34.** Use  $\int_0^1 \log x dx = 1$  in Example 5.1.2 to find  $\lim_{n \rightarrow \infty} \frac{\sqrt[n]{n!}}{n}$ . Note that for the proper integral, we cannot directly say that the Riemann sum converges to the integral.

**Exercise 5.35.** Let  $[x]$  be the biggest integer  $\leq x$ . Consider the series  $\sum \frac{(-1)^{[\sqrt{n}]}}{n^p}$ , with  $p > 0$ . Let

$$x_n = \sum_{k=n^2}^{(n+1)^2-1} \frac{1}{k^p} = \frac{1}{n^{2p}} + \frac{1}{(n^2+1)^p} + \cdots + \frac{1}{((n+1)^2-1)^p}.$$

1. Use Exercise 5.15 to prove that  $\sum \frac{(-1)^{[\sqrt{n}]}}{n^p}$  converges if and only if  $\sum (-1)^n x_n$  converges.
2. Estimate  $x_n$  and prove that  $x_{n+1} - x_n = (2 - 4p)n^{-2p} + o(n^{-2p})$ .
3. Prove that  $\sum \frac{(-1)^{[\sqrt{n}]}}{n^p}$  converges if and only if  $p > \frac{1}{2}$ .

The Dirichlet and Abel tests for improper integrals (Propositions 5.1.2 and 5.1.3) can also be extended.

**Proposition 5.2.3 (Dirichlet Test).** Suppose  $x_n$  is monotone and  $\lim_{n \rightarrow \infty} x_n = 0$ . Suppose the partial sum of  $\sum y_n$  is bounded. Then  $\sum x_n y_n$  converges.

**Proposition 5.2.4 (Abel Test).** Suppose  $x_n$  is monotone and bounded. Suppose  $\sum y_n$  converges. Then  $\sum x_n y_n$  converges.

*Proof.* The proof is the discrete version of the proof for the convergence of improper integrals. In fact, the discrete version of the integration by parts already appeared in the proof of Theorem 4.5.3, and is given by (see Figure 5.2.1, compare Figure 4.5.1)

$$\sum_{k=1}^n x_k y_k + \sum_{k=1}^{n-1} (x_{k+1} - x_k) s_k = x_n s_n, \quad s_n = \sum_{i=1}^n y_i, \quad y_i = s_i - s_{i-1}.$$

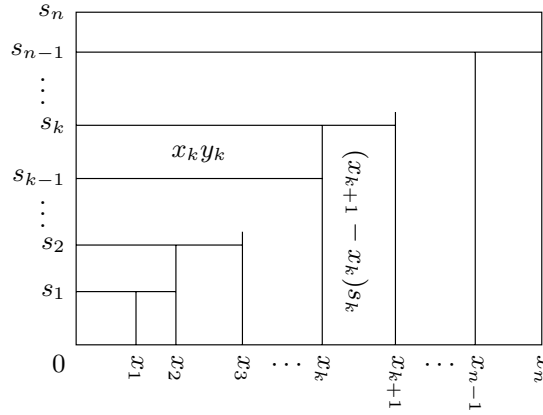
<sup>27</sup>Leonhard Paul Euler, born 1707 in Basel (Switzerland), died 1783 in St. Petersburg (Russia). Euler is one of the greatest mathematicians of all time. He made important discoveries in almost all areas of mathematics. Many theorems, quantities, and equations are named after Euler. He also introduced much of the modern mathematical terminology and notation, including  $f(x)$ ,  $e$ ,  $\Sigma$  (for summation),  $i$  (for  $\sqrt{-1}$ ), and modern notations for trigonometric functions.

<sup>28</sup>Lorenzo Mascheroni, born 1750 in Lombardo-Veneto (now Italy), died 1800 in Paris (France). The Euler-Mascheroni constant first appeared in a paper by Euler in 1735. Euler calculated the constant to 6 decimal places in 1734, and to 16 decimal places in 1736. Mascheroni calculated the constant to 20 decimal places in 1790.

Under the assumption of either test, we know  $s_n$  is bounded by a constant  $B$ ,  $\lim x_n$  converges, and  $\lim x_n s_n$  converges. Therefore the convergence of  $\sum x_n y_n$  is reduced to the convergence of  $\sum (x_{n+1} - x_n) s_n$ . Since  $x_n$  is monotone, the increments  $x_{k+1} - x_k$  do not change sign, and we have

$$\left| \sum_{k=m}^{n-1} (x_{k+1} - x_k) s_k \right| \leq \sum_{k=m}^{n-1} B |x_{k+1} - x_k| = B \left| \sum_{k=m}^{n-1} (x_{k+1} - x_k) \right| = B |x_n - x_m|.$$

Then the Cauchy criterion for the convergence of  $\lim x_n$  implies the Cauchy criterion for the convergence of  $\sum (x_{n+1} - x_n) s_n$ .  $\square$



**Figure 5.2.1.** The sum of vertical and horizontal strips is  $x_n s_n$ .

**Example 5.2.4.** The series  $\sum \frac{\sin na}{n}$  is the discrete version of Example 5.1.3. We may assume that  $a$  is not a multiple of  $\pi$  because otherwise the series is  $\sum 0 = 0$ . The partial sum of  $\sum \sin na$  is

$$\sum_{k=1}^n \sin ka = \frac{1}{2 \sin \frac{1}{2}a} \left[ \cos \frac{1}{2}a - \cos \left( n + \frac{1}{2} \right) a \right].$$

This is bounded by  $\frac{1}{\left| \sin \frac{1}{2}a \right|}$ . Moreover, the sequence  $\frac{1}{n}$  is decreasing and converges to 0.

By the Dirichlet test, the series converges.

To determine the nature of the convergence, we consider  $\sum \frac{|\sin na|}{n}$ . Note that although the series is similar to  $\int_1^{+\infty} \frac{|\sin x|}{x} dx$ , we cannot use the comparison test because  $\frac{|\sin x|}{x}$  is not a decreasing function.

We first assume  $0 < a < \pi$  and use the idea in Example 1.5.3 to find a strictly increasing sequence of natural numbers  $m_k$  satisfying

$$m_k a \in \left[ 2k\pi + \frac{a}{2}, 2k\pi + \pi - \frac{a}{2} \right].$$

Then

$$\sum_{n=1}^{m_k} \frac{|\sin na|}{n} \geq \sum_{i=1}^k \frac{|\sin m_i a|}{m_i a} \geq \sum_{i=1}^k \frac{\sin \frac{a}{2}}{2i\pi} = \frac{\sin \frac{a}{2}}{2\pi} \sum_{i=1}^k \frac{1}{i}.$$

Since  $\sum_{i=1}^{\infty} \frac{1}{i} = +\infty$ , the inequality above implies  $\sum_{n=1}^{\infty} \frac{|\sin na|}{n} = +\infty$ .

In general, if  $a$  is not a multiple of  $\pi$ , then we may find  $0 < b < \pi$ , such that  $a - b$  or  $a + b$  is a multiple of  $2\pi$ . By  $|\sin na| = |\sin nb|$ , the problem is reduced to the discussion above. We conclude that  $\sum \frac{\sin na}{n}$  conditionally converges as long as  $a$  is not a multiple of  $\pi$ .

**Exercise 5.36.** Derive the Abel test from the Dirichlet test.

**Exercise 5.37.** Derive the *Leibniz test* from the Dirichlet or Abel test: If  $x_n$  is decreasing and  $\lim_{n \rightarrow \infty} x_n = 0$ , then  $\sum (-1)^n x_n$  converges.

**Exercise 5.38.** Determine the convergence of  $\sum \frac{\sin na}{n^p}$ .

**Exercise 5.39.** Consider the series  $\sum \frac{\sin \sqrt{n}}{n^p}$ , with  $p > 0$ .

1. Prove that there is a constant  $b > 0$ , such that

$$\sum_{(2k - \frac{1}{4})\pi < \sqrt{n} < (2k + \frac{1}{4})\pi} \frac{\sin \sqrt{n}}{n^p} \geq \frac{b}{k^{2p-1}} \quad \text{for sufficiently big } k.$$

2. Use Exercise 4.103 to prove that there is a constant  $B > 0$ , such that

$$\left| \int_n^{n+1} \frac{\sin \sqrt{x}}{x^p} dx - \frac{1}{2} \left( \frac{\sin \sqrt{n}}{n^p} + \frac{\sin \sqrt{n+1}}{(n+1)^p} \right) \right| \leq \frac{B}{n^{p+\frac{1}{2}}}.$$

Then prove that, for  $p > \frac{1}{2}$ ,  $\sum \frac{\sin \sqrt{n}}{n^p}$  converges if and only if  $\int_1^{\infty} \frac{\sin \sqrt{x}}{x^p} dx$  converges.

3. Prove that the series absolutely converges if and only if  $p > 1$ , and conditionally converges if and only if  $\frac{1}{2} < p \leq 1$ .

## Rearrangement

A *rearrangement* of a series  $\sum x_n$  is  $\sum x_{k_n}$ , where  $n \rightarrow k_n$  is a one-to-one correspondence from the index set to itself (i.e., a rearrangement of the indices). The behaviour of rearranged series is directly related to whether the convergence of the series is absolute or conditional.

**Theorem 5.2.5.** *Any rearrangement of an absolutely convergent series is still absolutely convergent. Moreover, the sum is the same.*

*Proof.* Let  $s = \sum x_n$ . For any  $\epsilon > 0$ , there is a natural number  $N$ , such that  $\sum_{i=N+1}^{\infty} |x_i| < \epsilon$ . Let  $N' = \max\{i: k_i \leq N\}$ . Then  $\sum_{i=1}^{N'} x_{k_i}$  contains all the terms  $x_1, x_2, \dots, x_N$ . Therefore for  $n > N'$ , the difference  $\sum_{i=1}^n x_{k_i} - \sum_{i=1}^N x_i$  is a sum of some non-repeating terms in  $\sum_{i=N+1}^{\infty} x_i$ . This implies

$$\left| \sum_{i=1}^n x_{k_i} - \sum_{i=1}^N x_i \right| \leq \sum_{i=N+1}^{\infty} |x_i|,$$

and

$$\left| \sum_{i=1}^n x_{k_i} - s \right| \leq \left| \sum_{i=1}^n x_{k_i} - \sum_{i=1}^N x_i \right| + \left| \sum_{i=N+1}^{\infty} x_i \right| \leq 2 \sum_{i=N+1}^{\infty} |x_i| < 2\epsilon.$$

The absolute convergence of the rearrangement may be obtained by applying what was just proved to  $\sum |x_n|$ .  $\square$

**Theorem 5.2.6 (Riemann).** *A conditionally convergent series may be rearranged to have any number as the sum, or to become divergent.*

*Proof.* Suppose  $\sum x_n$  conditionally converges. Then  $\lim_{n \rightarrow \infty} x_n = 0$ . Let  $\sum x'_n$  and  $\sum x''_n$  be the series obtained by respectively taking only the non-negative terms and the negative terms. If  $\sum x'_n$  converges, then  $\sum x''_n = \sum x_n - \sum x'_n$  also converges. Therefore  $\sum |x_n| = \sum x'_n - \sum x''_n$  converges. Since  $\sum |x_n|$  is assumed to diverge, the contradiction shows that  $\sum x'_n$  diverges. Because  $x'_n \geq 0$ , we have  $\sum x'_n = +\infty$ . Similarly, we also have  $\sum x''_n = -\infty$ .

We will prove that the properties

$$\sum x'_n = +\infty, \quad \sum x''_n = -\infty, \quad \lim_{n \rightarrow \infty} x_n = 0$$

are enough for us to construct an arrangement with any number  $s$  as the limit. To simplify the notation, we introduce  $s_{(m,n]} = x_{m+1} + \dots + x_n$ . Then the partial sum  $s_{n_k} = s_{(0,n_k]} = s_{(0,n_1]} + s_{(n_1,n_2]} + \dots + s_{(n_{k-1},n_k]}$ . We use  $s', s''$  to denote the partial sums of  $\sum x'_n$  and  $\sum x''_n$ .

The idea is to keep adding positive terms until we are above  $s$ , and then keep adding negative terms until we are below  $s$ , and then adding positive terms and repeat the process. Specifically, by  $\sum x'_n = +\infty$ , there is  $m_1$ , such that

$$s'_{(0,m_1]} - x'_{m_1} = s'_{(0,m_1-1]} \leq s < s'_{(0,m_1]}.$$

By  $\sum x''_n = -\infty$ , there is  $n_1$ , such that

$$s'_{(0,m_1]} + s''_{(0,n_1]} - x''_{n_1} = s'_{(0,m_1]} + s''_{(0,n_1-1]} \geq s > s'_{(0,m_1]} + s''_{(0,n_1]}.$$

By  $\sum_{n > m_1} x'_n = +\infty$ , there is  $m_2 > m_1$ , such that

$$s'_{(0,m_1]} + s''_{(0,n_1]} + s'_{(m_1,m_2-1]} \leq s < s'_{(0,m_1]} + s''_{(0,n_1]} + s'_{(m_1,m_2]}.$$



Keep going, we get

$$\begin{aligned} & s'_{(0,m_1]} + s''_{(0,n_1]} + \cdots + s'_{(m_{k-1},m_k-1]} \\ &= s'_{(0,m_1]} + s''_{(0,n_1]} + \cdots + s'_{(m_{k-1},m_k]} - x'_{m_k} \\ &\leq s < s'_{(0,m_1]} + s''_{(0,n_1]} + \cdots + s'_{(m_{k-1},m_k]}, \end{aligned}$$

and

$$\begin{aligned} & s'_{(0,m_1]} + s''_{(0,n_1]} + \cdots + s'_{(m_{k-1},m_k]} + s''_{(n_{k-1},n_k-1]} \\ &= s'_{(0,m_1]} + s''_{(0,n_1]} + \cdots + s'_{(m_{k-1},m_k]} + s''_{(n_{k-1},n_k]} - x''_{n_k} \\ &\geq s > s'_{(0,m_1]} + s''_{(0,n_1]} + \cdots + s'_{(m_{k-1},m_k]} + s''_{(n_{k-1},n_k]}. \end{aligned}$$

By  $\lim_{n \rightarrow \infty} x_n = 0$ , the above implies that the special partial sums  $s'_{(0,m_1]} + s''_{(0,n_1]} + \cdots + s'_{(m_{k-1},m_k]}$  and  $s'_{(0,m_1]} + s''_{(0,n_1]} + \cdots + s'_{(m_{k-1},m_k]} + s''_{(n_{k-1},n_k]}$  of the rearranged series

$$x'_1 + \cdots + x'_{m_1} + x''_1 + \cdots + x''_{n_1} + x'_{m_1+1} + \cdots + x'_{m_2} + x''_{n_1+1} + \cdots + x''_{n_2} + \cdots$$

converge to  $s$ . Moreover, it is easy to see that the general partial sum of the rearranged series is sandwiched between two special partial sums. Therefore the general partial sum also converges to  $s$ .  $\square$

**Exercise 5.40.** Suppose  $x_n \geq 0$  and  $\sum x_n$  converges. Directly prove that, if  $\sum x_n$  converges, then any rearrangement  $\sum x_{k_n}$  converges, and  $\sum x_{k_n} \leq \sum x_n$ . Then further prove  $\sum x_{k_n} = \sum x_n$ .

**Exercise 5.41.** For any series  $\sum x_n$ , define two non-negative series  $\sum x'_n$  and  $\sum x''_n$  by

$$x'_n = \begin{cases} x_n, & \text{if } x_n \geq 0, \\ 0, & \text{if } x_n < 0; \end{cases} \quad x''_n = \begin{cases} 0, & \text{if } x_n > 0, \\ -x_n, & \text{if } x_n \leq 0. \end{cases}$$

Prove that  $\sum x_n$  absolutely converges if and only if both  $\sum x'_n$  and  $\sum x''_n$  converges. Moreover, prove that  $\sum x_n = \sum x'_n - \sum x''_n$  and  $\sum |x_n| = \sum x'_n + \sum x''_n$ .

**Exercise 5.42.** Use Exercises 5.40, 5.41, and the comparison test to give an alternative proof of Theorem 5.2.5.

**Exercise 5.43.** Rearrange the series  $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$  so that  $p$  positive terms are followed by  $q$  negative terms and the pattern repeated. Use Exercise 5.32 to show that the sum of the new series is  $\log 2 + \frac{1}{2} \log \frac{p}{q}$ . For any real number, expand the idea to construct a rearrangement to have the given number as the limit.

**Exercise 5.44.** Prove that if the rearrangement satisfies  $|k_n - n| < B$  for a constant  $B$ , then  $\sum x_{k_n}$  converges if and only if  $\sum x_n$  converges, and the sums are the same.

**Exercise 5.45.** Let  $0 < p \leq 1$  and let  $\sum \frac{(-1)^{k_n}}{k_n^p}$  be a rearrangement of  $\sum \frac{(-1)^n}{n^p}$ . Prove that if  $\lim_{n \rightarrow \infty} \frac{k_n - n}{n^p} = 0$ , then  $\sum \frac{(-1)^{k_n}}{k_n^p}$  converges and has the same sum as  $\sum \frac{(-1)^n}{n^p}$ .

**Exercise 5.46.** Suppose  $\sum x_n$  conditionally converges. Prove that for any  $a \leq b$ , there is a rearrangement, such that the partial sum  $s'_n$  satisfies  $\underline{\lim}_{n \rightarrow \infty} s'_n = a$ ,  $\overline{\lim}_{n \rightarrow \infty} s'_n = b$ . Moreover, prove that any number between  $a$  and  $b$  is the limit of a convergent subsequence of  $s'_n$ .

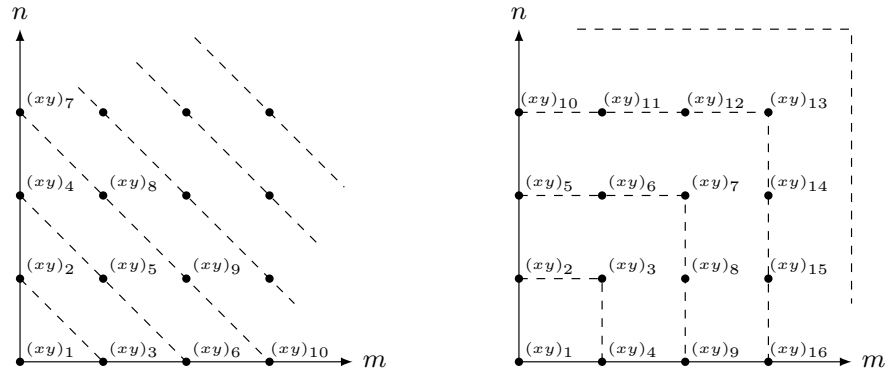
## Product of Series

The product of two series  $\sum x_n$  and  $\sum y_n$  involves the product  $x_m y_n$  for all  $m$  and  $n$ . In general, the *product series*  $\sum x_m y_n$  makes sense only after we arrange all the terms into a linear sequence  $\sum (xy)_k = \sum x_{m_k} y_{n_k}$ , which is given by a one-to-one correspondence  $(m_k, n_k): \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$ . For example, the following is the “diagonal arrangement”

$$\begin{aligned} \sum (xy)_k &= x_1 y_1 + x_1 y_2 + x_2 y_1 + \cdots \\ &\quad + x_1 y_{n-1} + x_2 y_{n-2} + \cdots + x_{n-1} y_1 + \cdots, \end{aligned}$$

and the following is the “square arrangement”

$$\begin{aligned} \sum (xy)_k &= x_1 y_1 + x_1 y_2 + x_2 y_2 + x_2 y_1 + \cdots \\ &\quad + x_1 y_n + x_2 y_n + \cdots + x_n y_{n-1} + x_n y_n + x_n y_{n-1} + \cdots + x_n y_1 + \cdots. \end{aligned}$$



**Figure 5.2.2.** Diagonal and square arrangements.

In view of Theorem 5.2.5, the following result shows that the arrangement does not matter if the series absolutely converge.

**Theorem 5.2.7.** Suppose  $\sum x_n$  and  $\sum y_n$  absolutely converge. Then  $\sum x_m y_n$  also absolutely converges, and  $\sum x_m y_n = (\sum x_m)(\sum y_n)$ .

*Proof.* Let  $s_n, t_n, S_n, T_n$  be the partial sums of  $\sum_{i=1}^{\infty} x_i, \sum_{i=1}^{\infty} y_i, \sum_{i=1}^{\infty} |x_i|, \sum_{i=1}^{\infty} |y_i|$ . Then we have convergent limits

$$\lim_{n \rightarrow \infty} s_n t_n = \left( \sum_{i=1}^{\infty} x_i \right) \left( \sum_{i=1}^{\infty} y_i \right), \quad \lim_{n \rightarrow \infty} S_n T_n = \left( \sum_{i=1}^{\infty} |x_i| \right) \left( \sum_{i=1}^{\infty} |y_i| \right).$$

Therefore for any  $\epsilon > 0$ , there is  $N$ , such that

$$\left| s_N t_N - \left( \sum_{i=1}^{\infty} x_i \right) \left( \sum_{i=1}^{\infty} y_i \right) \right| < \epsilon,$$

and

$$M > N \implies S_M T_M - S_N T_N < \epsilon.$$

Let  $\sum_{i=1}^{\infty} (xy)_i = \sum_{i=1}^{\infty} x_{m_i} y_{n_i}$  be any arrangement of the product series. Let  $K = \max\{i: m_i \leq N \text{ and } n_i \leq N\}$ . Then for any  $k > K$ ,  $\sum_{i=1}^k (xy)_i$  contains all the terms  $x_m y_n$  with  $1 \leq m, n \leq N$ , which are exactly the terms in  $s_N t_N = \sum_{1 \leq m, n \leq N} x_m y_n$ . Moreover, for each such  $k$ , the sum  $\sum_{i=1}^k (xy)_i$  is finite. Let  $M = \max\{\max\{m_i, n_i\}: 1 \leq i \leq k\}$ . Then all the terms in  $\sum_{i=1}^k (xy)_i$  are contained in  $s_M t_M = \sum_{1 \leq m, n \leq M} x_m y_n$ . Therefore the difference  $\sum_{i=1}^k (xy)_i - s_N t_N$  is a sum of some non-repeating terms in  $s_M t_M - s_N t_N$ . Since  $S_M S_M - S_N S_N$  is the sum of the absolute value of all the terms in  $s_M t_M - s_N t_N$ , we get

$$\begin{aligned} k > K &\implies \left| \sum_{i=1}^k (xy)_i - s_N t_N \right| \leq S_M S_M - S_N S_N < \epsilon \\ &\implies \left| \sum_{i=1}^k (xy)_i - \left( \sum_{i=1}^{\infty} x_i \right) \left( \sum_{i=1}^{\infty} y_i \right) \right| \\ &\leq \left| \sum_{i=1}^k (xy)_i - s_N t_N \right| + \left| s_N t_N - \left( \sum_{i=1}^{\infty} x_i \right) \left( \sum_{i=1}^{\infty} y_i \right) \right| < 2\epsilon. \end{aligned}$$

This proves that  $\lim_{k \rightarrow \infty} \sum_{i=1}^k (xy)_i = \left( \sum_{i=1}^{\infty} x_i \right) \left( \sum_{i=1}^{\infty} y_i \right)$ .

The absolute convergence of the product series may be obtained by applying what was just proved to  $\sum |x_m|$  and  $\sum |y_n|$ .  $\square$

**Example 5.2.5.** The geometric series  $\sum_{n=0}^{\infty} a^n$  absolutely converges to  $\frac{1}{1-a}$  for  $|a| < 1$ . The product of two copies of the geometric series is

$$\sum_{i,j \geq 0} a^i a^j = \sum_{n=0}^{\infty} \sum_{i+j=n} a^n = \sum_{n=0}^{\infty} (n+1) a^n.$$

Thus we conclude

$$1 + 2a + 3a^2 + \cdots + (n+1)a^n + \cdots = \frac{1}{(1-a)^2}.$$

**Exercise 5.47.** Suppose  $\sum x_n$  and  $\sum y_n$  absolutely converge. Directly prove that  $\sum x_n y_n$  absolutely converges. Then use Theorem 5.2.5 and a special arrangement to show that the sum of  $\sum x_n y_n$  is  $(\sum x_n)(\sum y_n)$ . This gives an alternative proof of Theorem 5.2.7.

**Exercise 5.48.** Suppose  $\sum x_n$  and  $\sum y_n$  converge (not necessarily absolutely). Prove that the square arrangement converges to  $(\sum x_n)(\sum y_n)$ .

**Exercise 5.49.** Suppose  $\sum (xy)_i$  is an arrangement of the product series, such that if  $i < j$ , then  $x_i y_k$  is arranged before  $x_j y_k$ .

1. Prove that the condition is equivalent to the partial sums of  $\sum (xy)_i$  are of the form  $s_{n_1} y_1 + s_{n_2} y_2 + \cdots + s_{n_k} y_k$ , where  $s_n$  are the partial sums of  $\sum x_n$ .
2. Prove that if  $\sum x_n$  converges and  $\sum y_n$  absolutely converges. Then  $\sum (xy)_i$  converges to  $(\sum x_n)(\sum y_n)$ .
3. Show that the square and diagonal arrangements satisfy the condition. On the other hand, show that the condition of absolute convergence in the second part is necessary by considering the diagonal arrangement of the product of  $\sum \frac{(-1)^n}{\sqrt{n}}$  with itself.

## 5.3 Uniform Convergence

The limit of a quantity happens when a variable approaches certain target. If the quantity has another parameter, then the “speed” of the convergence may generally depend on the parameter. If the speed is independent of the other parameter, then we say the convergence is *uniform*.

### Uniform Convergence of Two Variable Function

Suppose  $f(x, y)$  is a function of two variables  $x \in X \subset \mathbb{R}$  and  $y \in Y$ . We may take the limit with respect to the first variable and get a function

$$\lim_{x \rightarrow a} f(x, y) = g(y).$$

Strictly speaking, this means that, for any  $\epsilon > 0$  and  $y$ , there is  $\delta = \delta(\epsilon, y) > 0$  that may depend on  $\epsilon$  and  $y$ , such that

$$0 < |x - a| < \delta, x \in X \implies |f(x, y) - g(y)| < \epsilon.$$

The convergence is uniform if  $\delta$  is independent of  $y$ . In other words, for any  $\epsilon > 0$ , there is  $\delta = \delta(\epsilon) > 0$  (that may depend on  $\epsilon$  but is independent of  $y$ ), such that

$$0 < |x - a| < \delta, x \in X, y \in Y \implies |f(x, y) - g(y)| < \epsilon.$$

**Example 5.3.1.** Consider the function  $f(x, y) = \frac{y}{x+y}$  for  $x \in (0, +\infty)$  and  $y \in (0, 2]$ .

We have  $\lim_{x \rightarrow +\infty} f(x, y) = 0$  for each fixed  $y$ . Moreover, the following shows that the convergence is uniform

$$x > \frac{1}{2\epsilon} \implies |f(x, y) - 0| = \frac{y}{x+y} \leq \frac{2}{x} < \epsilon.$$

In fact, the convergence is uniform as long as the domain for  $y$  is bounded. On the other hand, if the domain for  $y$  is  $(0, +\infty)$ , then for  $\epsilon = \frac{1}{2}$  and any  $N$ , we may choose  $x = y = N + 1$  and get

$$x > N, y \in (0, +\infty), \text{ but } |f(x, y) - 0| = \frac{N+1}{2N+2} = \frac{1}{2}.$$

This shows that the convergence is not uniform.

**Example 5.3.2.** Consider

$$\lim_{x \rightarrow 0^+} \frac{\log(1+xy)}{x} = y, \quad y \geq 0.$$

We may use the linear approximation of  $\log(1+t)$  with Lagrangian remainder to get

$$\frac{\log(1+xy)}{x} - y = \frac{xy - \frac{1}{2(1+c)^2}(xy)^2}{x} - y = -\frac{1}{2(1+c)^2}x^2y, \quad 0 \leq c \leq xy.$$

We fix  $R > 0$  and take  $[0, R]$  as the domain of  $y$ . Then

$$0 < x < \frac{1}{R} \implies 0 \leq c \leq xy < 1 \implies \left| \frac{\log(1+xy)}{x} - y \right| \leq \frac{1}{2}x^2y < \frac{1}{2}x.$$

Therefore for any  $\frac{1}{R} > \epsilon > 0$ , we get

$$0 < x < \epsilon \implies \left| \frac{\log(1+xy)}{x} - y \right| < \frac{1}{2}x < \epsilon.$$

This shows the uniform convergence for  $0 \leq y \leq R$ .

Next we remove the bound on  $y$ . If the convergence is uniform for  $y \geq 0$ , then for  $\epsilon = 1$ , there is  $\delta > 0$ , such that

$$0 < x \leq \delta \implies \left| \frac{\log(1+xy)}{x} - y \right| < 1, \text{ for any } y \geq 0.$$

Taking  $x = \delta$  and letting  $y \rightarrow +\infty$ , the implication above gives

$$1 \geq \lim_{y \rightarrow +\infty} \left| \frac{\log(1+\delta y)}{\delta} - y \right| = \lim_{y \rightarrow +\infty} |y| \left| \frac{\log(1+\delta y)}{\delta y} - 1 \right| = +\infty.$$

The contradiction shows that the convergence is not uniform for  $y \geq 0$ . In fact, the same argument shows the non-uniformity for  $y \geq R$ .

The idea of using limit in  $y$  to argue against uniformity is a useful technique, and is related to the exchange of limits in Theorem 5.4.1. See Exercise 5.52.

We may also consider

$$\lim_{x \rightarrow 0^-} \frac{\log(1+xy)}{x} = y, \quad y < 0.$$

The convergence is uniform for  $-R \leq y \leq 0$  but is not uniform for  $y \leq -R$ .

**Example 5.3.3.** By applying the exponential function to the limit in Example 5.3.2, we get

$$\lim_{x \rightarrow 0^+} (1 + xy)^{\frac{1}{x}} = e^y, \quad y \geq 0.$$

In general, suppose  $\lim_{x \rightarrow a} f(x, y) = g(y)$  uniformly for  $y \in Y$ , and  $\varphi$  is a uniformly continuous function on the values of  $f(x, y)$  and  $g(y)$ . We claim that  $\lim_{x \rightarrow a} \varphi(f(x, y)) = \varphi(g(y))$  uniformly for  $y \in Y$ .

In fact, by the uniform continuity of  $\varphi$ , for any  $\epsilon > 0$ , there is  $\mu > 0$ , such that

$$|z - z'| < \mu \implies |\varphi(z) - \varphi(z')| < \epsilon.$$

Then by the uniformity of  $\lim_{x \rightarrow a} f(x, y) = g(y)$ , for  $\mu > 0$  obtained above, there is  $\delta > 0$ , such that

$$0 < |x| < \delta, y \in Y \implies |f(x, y) - g(y)| < \mu.$$

Combining two implications, we get

$$0 < |x| < \delta, y \in Y \implies |f(x, y) - g(y)| < \mu \implies |\varphi(f(x, y)) - \varphi(g(y))| < \epsilon.$$

This proves the uniformity of  $\lim_{x \rightarrow a} \varphi(f(x, y)) = \varphi(g(y))$ .

By Example 5.3.2, for any fixed  $R > 0$  and  $1 > 0$ , there is  $\delta > 0$ , such that

$$0 < x < \delta, 0 \leq y \leq R \implies \left| \frac{\log(1 + xy)}{x} - y \right| < 1 \implies \left| \frac{\log(1 + xy)}{x} \right| \leq R + 1.$$

Therefore for  $0 \leq y \leq R$  and sufficiently small  $x$ , the values of  $\frac{\log(1 + xy)}{x}$  and  $y$  lie in  $[-R - 1, R + 1]$ . Since the continuous function  $e^z$  is uniformly continuous on closed and bounded interval  $[-R - 1, R + 1]$ , the uniform convergence of  $\lim_{x \rightarrow 0^+} \frac{\log(1 + xy)}{x} = y$  for  $0 \leq y \leq R$  implies the uniform convergence of  $\lim_{x \rightarrow 0^+} (1 + xy)^{\frac{1}{x}} = e^y$  for  $0 \leq y \leq R$ .

The uniformity cannot be extended to unbounded  $y$ . If  $\lim_{x \rightarrow 0^+} (1 + xy)^{\frac{1}{x}} = e^y$  is uniform for  $y > 1$ , then by applying the function  $\log z$ , which is uniformly continuous for  $z \geq 1$ , we find that  $\lim_{x \rightarrow 0^+} \frac{\log(1 + xy)}{x} = y$  is uniform for  $y > 1$ . Since this is not true by Example 5.3.2, we conclude that  $\lim_{x \rightarrow 0^+} (1 + xy)^{\frac{1}{x}} = e^y$  is not uniform for  $y > 1$ .

Next we turn to

$$\lim_{x \rightarrow 0^-} (1 + xy)^{\frac{1}{x}} = e^y, \quad y \leq 0.$$

By the same argument, we find that the convergence is uniform for  $-R \leq y \leq 0$ . We will argue that the uniformity actually extends to  $-\infty$ .

For any  $\epsilon > 0$ , pick  $R > 0$  satisfying  $e^{-R} < \epsilon$ . We have

$$y < -R, x < 0 \implies 1 + xy > 1 - Rx \implies (1 + xy)^{\frac{1}{x}} < (1 - Rx)^{\frac{1}{x}}.$$

By  $\lim_{x \rightarrow 0^-} (1 - Rx)^{\frac{1}{x}} = e^{-R} < \epsilon$ , there is  $\delta_1 > 0$ , such that

$$-\delta_1 < x < 0 \implies (1 - Rx)^{\frac{1}{x}} < \epsilon.$$

Then

$$-\delta_1 < x < 0, y < -R \implies \left| (1 + xy)^{\frac{1}{x}} - e^y \right| \leq (1 + xy)^{\frac{1}{x}} + e^y < (1 - Rx)^{\frac{1}{x}} + e^{-R} < 2\epsilon.$$

On the other hand, the uniformity of  $\lim_{x \rightarrow 0^-} (1 + xy)^{\frac{1}{x}} = e^y$  for  $-R \leq y \leq 0$  implies that there is  $\delta_2 > 0$ , such that

$$-\delta_2 < x < 0, -R \leq y \leq 0 \implies \left| (1 + xy)^{\frac{1}{x}} - e^y \right| < \epsilon.$$

Combining the two implications, we get

$$-\min\{\delta_1, \delta_2\} < x < 0, y \leq 0 \implies \left| (1 + xy)^{\frac{1}{x}} - e^y \right| < 2\epsilon.$$

This completes the proof that  $\lim_{x \rightarrow 0^-} (1 + xy)^{\frac{1}{x}} = e^y$  is uniform for  $y \leq 0$ .

**Example 5.3.4.** We may change  $x$  in the limits in Example 5.3.3 to  $\frac{1}{x}$  and get

$$\lim_{x \rightarrow +\infty} \left(1 + \frac{y}{x}\right)^x = e^y \text{ for } y \geq 0, \quad \lim_{x \rightarrow -\infty} \left(1 + \frac{y}{x}\right)^x = e^y \text{ for } y \leq 0.$$

In general, we have a change of variable  $x = \varphi(z)$  satisfying  $\lim_{z \rightarrow b} \varphi(z) = a$ , and we may ask whether the uniformity of  $\lim_{x \rightarrow a} f(x, y) = g(y)$  implies the uniformity of  $\lim_{z \rightarrow b} f(\varphi(z), y) = g(y)$ . This is the uniform version of the composition rule, and is valid if one of the two additional conditions are satisfied.

- $z \neq c$  implies  $\varphi(z) \neq a$ .
- $g(y) = f(a, y)$  for all  $y \in Y$ . In other words, for each fixed  $y$ , the function  $f(x, y)$  of  $x$  is continuous at  $a$ .

We note that the conditions are only needed for the first limit to imply the second limit. The uniformity of the convergence is then transferred automatically. Applying the uniform composition rule to the limits in Example 5.3.3, the first limit is uniform for  $0 \leq y \leq R$  and is not uniform for  $y \geq R$ , and the second limit is uniform for  $y \leq 0$ .

**Example 5.3.5.** We may also change  $y$  in the limits in Example 5.3.3 and get

$$\lim_{x \rightarrow +\infty} \left(1 + \frac{\log y}{x}\right)^x = e^{\log y} = y, \quad y \geq 1.$$

In general, we have a change of variable  $y = \varphi(z): Z \rightarrow Y$ . Then it is easy to see that the uniformity of  $\lim_{x \rightarrow a} f(x, \varphi(z)) = g(\varphi(z))$  for  $z \in Z$  is equivalent to the uniformity of  $\lim_{x \rightarrow a} f(x, y) = g(y)$  for  $y \in \varphi(Z)$ . In particular, the uniformity of  $\lim_{x \rightarrow a} f(x, y) = g(y)$  for  $y \in Y$  implies the uniformity of  $\lim_{x \rightarrow a} f(x, \varphi(z)) = g(\varphi(z))$  for  $z \in Z$ .

In our case, we know  $\lim_{x \rightarrow +\infty} \left(1 + \frac{\log y}{x}\right)^x = y$  is uniform for  $1 \leq y \leq R$  and is not uniform for  $y \geq R$ .

**Exercise 5.50.** Prove that  $\lim_{x \rightarrow a} f(x, y) = g(y)$  uniformly on  $Y$  if and only if

$$\lim_{x \rightarrow a} \sup_Y |f(x, y) - g(y)| = 0.$$

**Exercise 5.51.** Prove that  $\lim_{x \rightarrow a} f(x, y) = g(y)$  uniformly on  $Y = Y_1 \cup Y_2$  if and only if it converges uniformly on  $Y_1$  and on  $Y_2$ . Extend the statement to finite union. Can you extend to infinite union?

**Exercise 5.52.** Suppose  $\lim_{x \rightarrow a} f(x, y) = g(y)$  for  $y \in (b, c)$ . Suppose there is  $C > 0$ , such that for each fixed  $x$ , we have  $|f(x, y) - g(y)| \geq C$  for  $y$  sufficiently close to  $b$ . Prove that the convergence of is not uniform on  $(b, r)$  for any  $b < r < c$ .

Note that the criterion  $|f(x, y) - g(y)| \geq C$  is satisfied if  $\lim_{y \rightarrow b+} |f(x, y) - g(y)|$  converges to a number  $> C$  or diverges to  $+\infty$ .

**Exercise 5.53.** Determine the uniform convergence (the answer may depend on the domain for  $y$ ).

1.  $\lim_{x \rightarrow \infty} \frac{1}{xy} = 0$ .
2.  $\lim_{x \rightarrow 0} \frac{\sin xy}{x} = y$ .
3.  $\lim_{x \rightarrow 0+} \sqrt{x+y} = \sqrt{y}$ .
4.  $\lim_{x \rightarrow 0} (1+xy)^{\frac{1}{x}} = e^y$ .
5.  $\lim_{x \rightarrow 0} y^x = 1$ .
6.  $\lim_{x \rightarrow 0} \frac{y^x - 1}{x} = \log y$ .

**Exercise 5.54.** Determine the uniform convergence of the definition of derivatives.

1.  $\lim_{x \rightarrow y} \frac{\sin x - \sin y}{x - y} = \cos y$ .
2.  $\lim_{x \rightarrow y} \frac{x^p - y^p}{x - y} = py^{p-1}, y > 0$ .

**Exercise 5.55.** Define the uniform convergence of a double sequence  $\lim_{n \rightarrow \infty} x_{m,n} = y_m$ . The determine the uniformity of the convergence.

1.  $\lim_{n \rightarrow \infty} \frac{1}{m+n} = 0$ .
2.  $\lim_{n \rightarrow \infty} \frac{m}{m+n} = 0$ .
3.  $\lim_{n \rightarrow \infty} \frac{1}{n^m} = 0$ .
4.  $\lim_{n \rightarrow \infty} \frac{m}{n} = 0$ .
5.  $\lim_{n \rightarrow \infty} \sqrt[n]{m} = 1$ .

**Exercise 5.56.** Suppose  $f$  is continuous on an open interval containing a bounded and closed interval  $[a, b]$ . Prove that  $\lim_{x \rightarrow 0} f(x+y) = f(y)$  uniformly on  $[a, b]$ .

**Exercise 5.57.** Suppose  $f$  is a function on an interval  $[a, b]$ . Extend  $f$  to a function on  $\mathbb{R}$  by setting  $f(x) = f(a)$  for  $x < a$  and  $f(x) = f(b)$  for  $x > b$ . Prove that  $f$  is uniformly continuous on  $[a, b]$  if and only if  $\lim_{x \rightarrow 0} f(x+y) = f(y)$  uniformly on  $[a, b]$ .

**Exercise 5.58.** Suppose  $f$  has continuous derivative on an open interval containing  $[a, b]$ . Prove that  $\lim_{x \rightarrow 0} \frac{f(x+y) - f(y)}{x} = f'(y)$  uniformly on  $[a, b]$ .

**Exercise 5.59.** Suppose  $f(x)$  is continuous on an open interval containing  $[a, b]$ . Prove that the Fundamental Theorem of Calculus

$$\lim_{x \rightarrow 0} \frac{1}{x} \int_y^{x+y} f(t) dt = f(y)$$

is uniform on  $[a, b]$ .



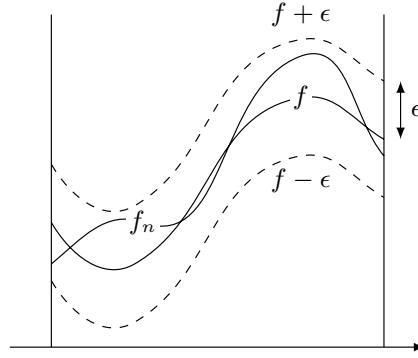
### Uniform Convergence of Sequence and Series of Functions

A sequence of functions  $f_n(x)$  uniformly converges to a function  $f(x)$  on domain  $X$  if for any  $\epsilon > 0$ , there is  $N$ , such that

$$n > N, x \in X \implies |f_n(x) - f(x)| < \epsilon.$$

Here  $n$  and  $x$  play the role of  $x$  and  $y$  in the uniform convergence of  $\lim_{x \rightarrow a} f(x, y)$ . As suggested by Examples 5.3.1 and 5.3.2, the uniformity of the convergence may depend on the domain for  $x$ .

Figure 5.3.1 shows that the uniform convergence of a sequence of functions means that the graph of  $f_n$  lies in the  $\epsilon$ -strip around  $f$  for sufficiently big  $n$ .



**Figure 5.3.1.** Uniform convergence of sequence of functions.

The partial sum of a series of functions  $\sum u_n(x)$  is a sequence of functions  $s_n(x) = u_1(x) + u_2(x) + \cdots + u_n(x)$ . The uniform convergence of the series is the uniform convergence of the sequence  $s_n(x)$ .

**Example 5.3.6.** For the sequence  $x^n$ , we have

$$\lim_{n \rightarrow \infty} x^n = \begin{cases} 0, & \text{if } x \in (-1, 1), \\ 1, & \text{if } x = 1, \\ \text{diverge,} & \text{otherwise.} \end{cases}$$

Denote by  $f(x)$  the limit function on the right, defined on the interval  $(-1, 1]$ .

If we take the domain of  $x$  to be the biggest  $X = (-1, 1]$ , then the convergence is not uniform. Specifically, for any  $N \in \mathbb{N}$ , by  $\lim_{x \rightarrow 1} x^{N+1} = 1$ , we can always find  $x$  very close to the left of 1, such that  $x^{N+1} > \frac{1}{2}$ . By  $f(x) = 0$ , this shows that

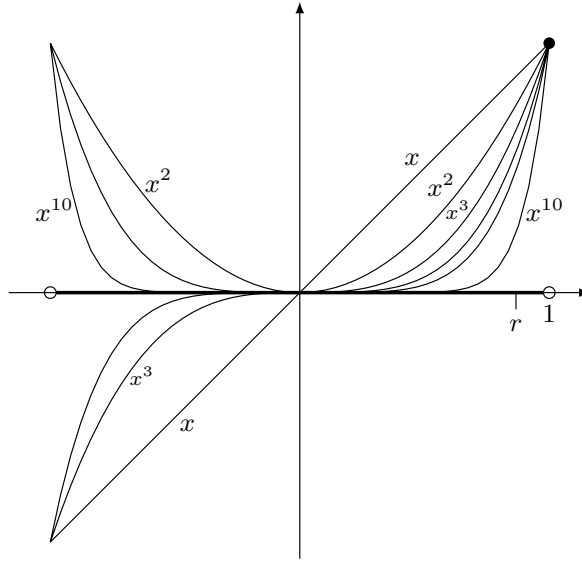
$$n = N + 1 > N, x \in (-1, 1], \text{ but } |x^n - f(x)| = x^{N+1} > \frac{1}{2}.$$

Therefore the uniform convergence fails for  $\epsilon = \frac{1}{2}$ . In fact, the argument shows that the convergence is not uniform on  $X = (r, 1)$  or on  $X = (-1, -r)$  for any  $0 < r < 1$ .

If we take  $X = [-r, r]$  for any fixed  $0 < r < 1$ , so that  $X$  is of some distance away from  $\pm 1$ , then the convergence is uniform on  $X$ . Specifically, by  $\lim_{n \rightarrow \infty} r^n = 0$ , for any  $\epsilon > 0$ , there is  $N$ , such that  $n > N$  implies  $r^n < \epsilon$ . Then

$$n > N, x \in [-r, r] \implies |x^n - f(x)| = |x|^n \leq r^n < \epsilon.$$

This verifies the uniformity of the convergence.



**Figure 5.3.2.**  $\lim_{n \rightarrow \infty} x^n$ .

**Example 5.3.7.** By Example 5.2.1, the partial sum of the series  $\sum_{n=0}^{\infty} x^n$  is  $\frac{1-x^{n+1}}{1-x}$ , the sum is  $\frac{1}{1-x}$  for  $x \in (-1, 1)$ , and

$$\left| \frac{1-x^{n+1}}{1-x} - \frac{1}{1-x} \right| = \frac{|x^{n+1}|}{|1-x|}.$$

For any fixed  $0 < r < 1$ , we have  $\lim_{n \rightarrow \infty} \frac{r^{n+1}}{1-r} = 0$ . In other words, for any  $\epsilon > 0$ , there is  $N = N(\epsilon, r)$ , such that  $n > N$  implies  $\frac{r^{n+1}}{1-r} < \epsilon$ . Then

$$n > N, |x| \leq r \implies \left| \frac{1-x^{n+1}}{1-x} - \frac{1}{1-x} \right| = \frac{|x^{n+1}|}{|1-x|} \leq \frac{r^{n+1}}{1-r} < \epsilon.$$

This shows that  $\sum x^n$  uniformly converges for  $|x| \leq r$ .

On the other hand, for any  $N$ , we may fix any natural number  $n > N$ . Then by

$$\lim_{x \rightarrow 1^-} \left| \frac{1-x^{n+1}}{1-x} - \frac{1}{1-x} \right| = \lim_{x \rightarrow 1^-} \frac{|x^{n+1}|}{1-x} = +\infty > 1,$$

we can find  $x$  satisfying

$$n > N, \quad 1 > x > r, \quad \text{but} \quad \left| \frac{1 - x^{n+1}}{1 - x} - \frac{1}{1 - x} \right| > 1.$$

This shows the failure of the uniform convergence on  $(r, 1)$  for  $\epsilon = 1$ . Similarly, by

$$\lim_{x \rightarrow (-1)^+} \frac{|x^{n+1}|}{1 - x} = \frac{1}{2} > \frac{1}{3},$$

we can show the failure of the uniform convergence on  $(-1, -r)$  for  $\epsilon = \frac{1}{3}$ .

The idea of using limit in  $x$  to argue against the uniformity is summarised in Exercise 5.62.

**Example 5.3.8.** We have  $\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$  similar to the limits of two variable function in Example 5.3.4. By the same argument, the sequence of functions uniformly converges for  $|x| \leq R$ . Moreover, by  $\lim_{x \rightarrow \infty} \left| \left(1 + \frac{x}{n}\right)^n - e^x \right| = +\infty$ , the convergence of the sequence of functions is not uniform for  $x \geq R$  or for  $x \leq -R$ .

**Example 5.3.9.** The Taylor series of  $e^x$  is  $\sum \frac{1}{n!} x^n$ . By the Lagrange remainder in Proposition 3.4.3, we have

$$\left| \sum_{k=0}^n \frac{1}{k!} x^k - e^x \right| = \frac{e^c |x|^{n+1}}{(n+1)!} \leq \frac{e^{|x|} |x|^{n+1}}{(n+1)!}, \quad c \text{ between } 0 \text{ and } x.$$

For any fixed  $R > 0$ , we have  $\lim_{n \rightarrow \infty} \frac{e^R R^{n+1}}{(n+1)!} = 0$ . Therefore for any  $\epsilon > 0$ , there is  $N = N(\epsilon, R)$ , such that  $n > N$  implying  $\frac{e^R R^{n+1}}{(n+1)!} < \epsilon$ . Then

$$n > N, \quad |x| \leq R \implies \left| \sum_{k=0}^n \frac{1}{k!} x^k - e^x \right| \leq \frac{e^{|x|} |x|^{n+1}}{(n+1)!} \leq \frac{e^R R^{n+1}}{(n+1)!} < \epsilon.$$

This shows that Taylor series converges to  $e^x$  uniformly for  $|x| \leq R$ .

On the other hand, for fixed  $n$ , we have

$$\lim_{x \rightarrow \infty} \left| \sum_{k=0}^n \frac{1}{k!} x^k - e^x \right| = +\infty.$$

By the same reason as in Example 5.3.7, the convergence of the Taylor series is not uniform for  $x < -R$  or  $x > R$ .

**Exercise 5.60.** Prove that  $f(x, y)$  uniformly converges to  $g(y)$  as  $x \rightarrow a$  if and only if  $f(x_n, y)$  uniformly converges to  $g(y)$  for any sequence  $x_n \neq a$  converging to  $a$ .

**Exercise 5.61.** Prove that if  $f_n$  and  $g_n$  uniformly converges, then  $af_n + bg_n$  uniformly converges. What about product, composition, maximum, etc, of uniformly convergent sequences of functions?

**Exercise 5.62.** Suppose  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$  for  $x$  near  $a$ . Suppose there is  $C > 0$ , such that for each fixed  $n$ , we have  $|f_n(x) - f(x)| \geq C$  for  $x$  sufficiently close to  $a$ . Prove that the convergence of  $f_n$  is not uniform for  $0 < |x - a| < \delta$  for any  $\delta > 0$ .

Note that the criterion  $|f_n(x) - f(x)| \geq c$  is satisfied if  $\lim_{x \rightarrow a} |f_n(x) - f(x)|$  converges to a number  $> c$  or diverges to  $+\infty$ .

**Exercise 5.63.** Suppose  $f(x)$  is integrable on  $[a, b + 1]$ . Prove that the sequence  $f_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} f\left(x + \frac{i}{n}\right)$  uniformly converges to  $\int_x^{x+1} f(t)dt$  on  $[a, b]$ .

**Exercise 5.64.** Determine the uniform convergence of sequences of functions (the answer may depend on the domain for  $x$ ).

- |                               |                          |  |
|-------------------------------|--------------------------|--|
| 1. $x^{\frac{1}{n}}$ .        | 5. $\frac{x}{n^x}$ .     | 9. $\left(x + \frac{1}{n}\right)^p$ .    |
| 2. $n(x^{\frac{1}{n}} - 1)$ . | 6. $\frac{\sin nx}{n}$ . | 10. $\left(1 + \frac{x}{n}\right)^p$ .   |
| 3. $\frac{1}{n+x}$ .          | 7. $\sin \frac{x}{n}$ .  | 11. $\log\left(1 + \frac{x}{n}\right)$ . |
| 4. $\frac{1}{nx+1}$ .         | 8. $\sqrt[n]{1+x^n}$ .   | 12. $\left(1 + \frac{x}{n}\right)^n$ .   |

## Uniform Convergence Test

The test for uniform convergence of sequence and series of functions starts with the Cauchy criterion.

**Proposition 5.3.1 (Cauchy Criterion).** *A sequence  $f_n(x)$  uniformly converges on  $X$  if and only if for any  $\epsilon > 0$ , there is  $N$ , such that*

$$m, n > N, x \in X \implies |f_m(x) - f_n(x)| < \epsilon.$$

*Proof.* Suppose  $f_n(x)$  uniformly converges to  $f(x)$  on  $X$ . Then for any  $\epsilon > 0$ , there is  $N$ , such that

$$m, n > N, x \in X \implies |f_n(x) - f(x)| < \epsilon.$$

This implies

$$m, n > N, x \in X \implies |f_m(x) - f_n(x)| \leq |f_m(x) - f(x)| + |f_n(x) - f(x)| < 2\epsilon.$$

which verifies the Cauchy criterion,

Conversely, suppose the uniform Cauchy criterion is satisfied. Then for each fixed  $x \in X$ , the criterion shows that  $f_n(x)$  is a Cauchy sequence, and therefore  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$  converges. To see that the uniformity of the convergence, we apply the uniform Cauchy criterion again. For any  $\epsilon > 0$ , there is  $N$ , such that

$$m, n > N, x \in X \implies |f_m(x) - f_n(x)| < \epsilon.$$

Now for each fixed  $n > N$ , we let  $m \rightarrow \infty$  and get

$$n > N, x \in X \implies |f(x) - f_n(x)| \leq \epsilon.$$

This shows that the convergence is uniform on  $X$ .  $\square$

The uniform Cauchy criterion for the series  $\sum u_n(x)$  is that, for any  $\epsilon > 0$ , there is  $N$ , such that

$$n \geq m > N, x \in X \implies |u_m(x) + u_{m+1}(x) + \cdots + u_n(x)| < \epsilon.$$

The special case  $m = n$  means that, if  $\sum u_n(x)$  uniformly converges, then  $u_n(x)$  uniformly converges to 0. For example, since the convergence of  $x^n$  is not uniform on  $(r, 1)$  (Example 5.3.6), the convergence of  $\sum x^n$  is also not uniform on  $(r, 1)$  (Example 5.3.7).

We used the Cauchy criterion to derive various tests for the convergence of series of numbers. Using Proposition 5.3.1, we may extend all the tests to the uniform convergence of series of functions. We say a sequence of functions  $f_n$  is *uniformly bounded* on  $X$  if there is  $B$ , such that  $|f_n(x)| < B$  for all  $n$  and  $x \in X$ .

**Proposition 5.3.2 (Comparison Test).** *Suppose  $|u_n(x)| \leq v_n(x)$ . If  $\sum v_n(x)$  uniformly converges, then  $\sum u_n(x)$  uniformly converges.*

**Proposition 5.3.3 (Dirichlet Test).** *Suppose  $u_n(x)$  is a monotone sequence for each  $x$ , and uniformly converges to 0. Suppose the partial sums of  $\sum v_n(x)$  are uniformly bounded. Then  $\sum u_n(x)v_n(x)$  uniformly converges.*

**Proposition 5.3.4 (Abel Test).** *Suppose  $u_n(x)$  is a monotone sequence for each  $x$ , and is uniformly bounded. Suppose  $\sum v_n(x)$  uniformly converges. Then  $\sum u_n(x)v_n(x)$  uniformly converges.*

The proof of the propositions are left as exercises. We note that in the monotone condition,  $u_n(x)$  is allowed to be increasing for some  $x$  and decreasing for some other  $x$ .

**Example 5.3.10.** We have  $\left| \frac{(-1)^n}{n^2 + x^2} \right| \leq \frac{1}{n^2}$ . By considering  $\frac{1}{n^2}$  as constant functions, the series  $\sum \frac{1}{n^2}$  uniformly converges. Then by the comparison test,  $\sum \frac{(-1)^n}{n^2 + x^2}$  also uniformly converges.

**Example 5.3.11.** For any  $0 < r < 1$ , we have  $\left| \frac{(-1)^n}{n} x^n \right| \leq r^n$  for  $|x| \leq r$ . By the (uniform) convergence of (constant function series)  $\sum r^n$  and the comparison test,  $\sum \frac{(-1)^n}{n} x^n$  uniformly converges for  $|x| \leq r$ .

We note that  $\frac{1}{n}$  is decreasing and (uniformly) converges to 0. Moreover, the partial sum of  $\sum (-1)^n x^n$  is uniformly bounded on  $[0, 1]$

$$\left| \sum_{k=0}^n (-1)^k x^k \right| = \frac{1 - (-x)^{n+1}}{1 + x} \leq 2.$$

By the Dirichlet test, therefore, the series  $\sum \frac{(-1)^n}{n} x^n$  uniformly converges on  $[0, 1]$ . Combined with  $[-r, r]$ , the series uniformly converges on  $[-r, 1]$ .

An alternative way of showing the uniform convergence of  $\sum \frac{(-1)^n}{n} x^n$  on  $[0, 1]$  is by using the uniform version of the Leibniz test in Exercise 5.67.

At  $x = -1$ , the series is the harmonic series in Example 1.5.2 and therefore diverges. Therefore we do not expect the convergence to be uniform on  $(-1, -r)$ . We consider the sum of the  $(n+1)$ -st term to the  $2n$ -th term (similar to Example 1.5.2) and take the limit as  $x \rightarrow (-1)^+$  (because this is where the trouble is)

$$\lim_{x \rightarrow (-1)^+} \left| \sum_{k=n+1}^{2n} \frac{(-1)^k}{k} (-x)^k \right| = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} > \frac{1}{2}.$$

For any  $N$ , fix a natural number  $n > N$ . The limit above implies that there is  $x \in (-1, -r)$ , such that

$$n, 2n > N, \quad x \in (-1, -r), \quad \text{but} \quad \left| \sum_{k=n+1}^{2n} \frac{(-1)^k}{k} (-x)^k \right| > \frac{1}{2}.$$

This shows that the Cauchy criterion for the uniform convergence fails on  $(-1, -r)$  for  $\epsilon = \frac{1}{2}$ .

**Example 5.3.12.** By Example 5.2.4, the series  $\sum \frac{\sin nx}{n}$  converges for all  $x$ . The series  $\frac{1}{n}$  is decreasing and uniformly converges to 0. Moreover, the calculation in Example 5.2.4 shows that the partial sum of  $\sum \sin nx$  is uniformly bounded on  $[r, \pi - r]$  for any  $0 < r < \frac{\pi}{2}$ . By the Dirichlet test, therefore, the series uniformly converges on  $[r, \pi - r]$  (and on  $[m\pi + r, (m+1)\pi - r]$ ,  $m \in \mathbb{Z}$ ).

Next we show that the uniform convergence cannot be extended to  $(0, r)$ . By  $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$ , we have  $\frac{\sin x}{x} > \frac{1}{2}$  on some interval  $(0, b)$ . (In fact,  $\frac{\sin x}{x} > \frac{2}{\pi}$  for  $0 < x < \frac{\pi}{2}$ .) For any  $N$ , fix any natural number satisfying  $n > N$  and  $n > \frac{b}{2r}$ . Then  $x = \frac{b}{2n}$  satisfies  $x \in (0, r)$ ,  $kx \in (0, b)$  for all  $n < k \leq 2n$ , and

$$\sum_{k=n+1}^{2n} \frac{\sin kx}{k} = x \sum_{k=n+1}^{2n} \frac{\sin kx}{kx} > \sum_{k=n+1}^{2n} \frac{1}{2} x = \frac{nx}{2} = \frac{b}{4}.$$

This shows that the Cauchy criterion for the uniform convergence of  $\sum \frac{\sin nx}{n}$  on  $(0, r)$  fails for  $\epsilon = \frac{b}{4}$ .

**Exercise 5.65.** Prove Propositions 5.3.2, 5.3.3, 5.3.4.

**Exercise 5.66.** State and prove the Cauchy criterion for the uniform convergences of two variable function.

**Exercise 5.67.** State and prove the uniform convergence version of the Leibniz test.

**Exercise 5.68.** Suppose  $\sum f_n(x)^2$  uniformly converges. Prove that  $\sum \frac{f_n(x)}{n^p}$  also uniformly converges for any  $p > \frac{1}{2}$ .

**Exercise 5.69.** Determine the intervals on which the series uniformly converge.

- |   |                                 |  |
|---|---------------------------------|--|
| 1. $\sum x^n e^{-nx}$ .                       | 5. $\sum \frac{1}{x + a^n}$ .   | 8. $\sum \frac{\sin^3 nx}{n^p}$ .                                    |
| 2. $\sum n^x x^n$ .                           | 6. $\sum \frac{x^n}{1 - x^n}$ . | 9. $\sum \frac{\sin nx}{n^p (\log n)^q}$ .                           |
| 3. $\sum \left( \frac{x(x+n)}{n} \right)^n$ . | 7. $\sum \frac{\cos nx}{n^p}$ . | 10. $\sum \left  e^x - \left( 1 + \frac{x}{n} \right)^n \right ^p$ . |
| 4. $\sum \frac{1}{n^p + x^p}$ .               |                                 |  |

**Exercise 5.70.** Show that  $\sum (-1)^n x^n (1 - x)$  uniformly converges on  $[0, 1]$  and absolutely converges for each  $x \in [0, 1]$ . However, the convergence of the absolute value series  $\sum |(-1)^n x^n (1 - x)|$  is not uniform on  $[0, 1]$ .

## 5.4 Exchange of Limits

A fundamental consequence of the uniform convergence is the exchange of limits. The following result is stated for the limit of sequence of functions.

**Theorem 5.4.1.** Suppose  $f_n(x)$  uniformly converges to  $f(x)$  for  $x \neq a$ . Suppose  $\lim_{x \rightarrow a} f_n(x) = l_n$  converges for each  $n$ . Then both  $\lim_{n \rightarrow \infty} l_n$  and  $\lim_{x \rightarrow a} f(x)$  converge and are equal.

The conclusion is the exchange of two limits

$$\lim_{n \rightarrow \infty} \lim_{x \rightarrow a} f_n(x) = \lim_{n \rightarrow \infty} l_n = \lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} \lim_{n \rightarrow \infty} f_n(x).$$

In other words, the two *repeated limits* are equal.

*Proof.* For any  $\epsilon > 0$ , there is  $N$ , such that

$$x \neq a \text{ and } m, n > N \implies |f_m(x) - f_n(x)| < \epsilon. \quad (5.4.1)$$

We take  $\lim_{x \rightarrow a}$  of the right side and get

$$m, n > N \implies |l_m - l_n| \leq \epsilon.$$

Therefore  $l_n$  is a Cauchy sequence and converges to a limit  $l$ . Moreover, we take  $m \rightarrow \infty$  in the implication above and get

$$n > N \implies |l - l_n| \leq \epsilon.$$

On the other hand, we take  $\lim_{m \rightarrow \infty}$  of the right side of (5.4.1) and get

$$x \neq a \text{ and } n > N \implies |f(x) - f_n(x)| \leq \epsilon.$$

Fix one  $n > N$ . Since  $\lim_{x \rightarrow a} f_n(x) = l_n$ , there is  $\delta > 0$ , such that

$$0 < |x - a| < \delta \implies |f_n(x) - l_n| < \epsilon.$$

Then for the fixed choice of  $n$ , we have

$$0 < |x - a| < \delta \implies |f(x) - l| \leq |f(x) - f_n(x)| + |f_n(x) - l_n| + |l - l_n| < 3\epsilon.$$

This proves that  $\lim_{x \rightarrow a} f(x) = l$ . □

**Example 5.4.1.** The convergence in Example 5.3.6 is uniform for  $|x| \leq r < 1$ . This suggests the exchange of limits at  $a \in (-r, r)$

$$\lim_{n \rightarrow \infty} \lim_{x \rightarrow a} x^n = \lim_{n \rightarrow \infty} a^n = 0, \quad \lim_{x \rightarrow a} \lim_{n \rightarrow \infty} x^n = \lim_{x \rightarrow a} 0 = 0.$$

Since  $r < 1$  is arbitrary, we have exchange of limits everywhere on  $(-1, 1)$ .

On the other hand, the convergence is not uniform on  $(0, 1)$ . The two repeated limits are not equal at  $1^-$

$$\lim_{n \rightarrow \infty} \lim_{x \rightarrow 1^-} x^n = \lim_{n \rightarrow \infty} 1 = 1, \quad \lim_{x \rightarrow 1^-} \lim_{n \rightarrow \infty} x^n = \lim_{x \rightarrow 1^-} 0 = 0.$$

**Example 5.4.2.** For the case  $f_n(x)$  is the partial sum of a series of functions  $\sum u_n(x)$ , Theorem 5.4.1 says that, if  $\sum u_n(x)$  uniformly converges, then the limit and the sum can be exchanged

$$\sum \lim_{x \rightarrow a} u_n(x) = \lim_{x \rightarrow a} \sum u_n(x).$$

The series  $\sum_{n=0}^{\infty} x^n$  in Example 5.3.7 uniformly converges in  $[-r, r]$  for any  $0 < r < 1$ . Since any  $a$  satisfying  $|a| < 1$  lies inside such interval, we expect the exchange of  $\lim_{x \rightarrow a}$  and the sum

$$\sum_{n=0}^{\infty} \lim_{x \rightarrow a} x^n = \sum_{n=0}^{\infty} a^n = \frac{1}{1-a}, \quad \lim_{x \rightarrow a} \sum_{n=0}^{\infty} x^n = \lim_{x \rightarrow a} \frac{1}{1-x} = \frac{1}{1-a}.$$

On the other hand, the series does not converge uniformly on  $(-1, 0)$ , and we have

$$\sum_{n=0}^{\infty} \lim_{x \rightarrow (-1)^+} x^n = \sum_{n=0}^{\infty} (-1)^n \text{ diverges,} \quad \lim_{x \rightarrow (-1)^+} \sum_{n=0}^{\infty} x^n = \lim_{x \rightarrow (-1)^+} \frac{1}{1-x} = \frac{1}{2}.$$

**Example 5.4.3.** The convergence in Example 5.3.1 is not uniform on  $(R, +\infty)$ . The two repeated limits at  $+\infty$  are not equal

$$\lim_{y \rightarrow +\infty} \lim_{x \rightarrow +\infty} \frac{y}{x+y} = \lim_{y \rightarrow +\infty} 0 = 0, \quad \lim_{x \rightarrow +\infty} \lim_{y \rightarrow +\infty} \frac{y}{x+y} = \lim_{x \rightarrow +\infty} 1 = 1.$$

**Exercise 5.71.** Determine whether two limits exchange for two variable functions in Exercise 5.53 and 5.54, double sequences in Exercise 5.55, the sequences in Exercise 5.64, and series in Exercise 5.69.



### Uniform Convergence and Continuity

Suppose  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$  uniformly on  $X$ . Suppose  $f_n$  are continuous at  $a \in X$ . Then by Theorem 5.4.1, we have

$$\begin{aligned}
 \lim_{x \rightarrow a} f(x) &= \lim_{x \rightarrow a} \lim_{n \rightarrow \infty} f_n(x) && \text{(definition of } f) \\
 &= \lim_{n \rightarrow \infty} \lim_{x \rightarrow a} f_n(x) && \text{(exchange of limit)} \\
 &= \lim_{n \rightarrow \infty} f_n(a) && \text{(continuity of } f_n \text{ at } a) \\
 &= f(a). && \text{(definition of } f)
 \end{aligned}$$

This shows that the uniform limit of continuous functions is continuous.

**Proposition 5.4.2.** *Suppose  $f_n(x)$  uniformly converges to  $f(x)$  for  $x \in X$ . If  $f_n$  are continuous at  $a \in X$ , then  $f$  is also continuous at  $a \in X$ .*

The following is one converse of the proposition.

**Proposition 5.4.3 (Dini's Theorem).** *Suppose  $f_n(x)$  converges to  $f(x)$  on a bounded and closed interval  $I$ . Suppose  $f_n(x)$  is monotone in  $n$  for each fixed  $x$ . If  $f_n(x)$  and  $f(x)$  are continuous on  $I$ , then the convergence of  $f_n$  to  $f$  is uniform on  $I$ .*

Note that we allow  $f_n(x)$  to be increasing for some  $x$  and decreasing for some other  $x$ .

*Proof.* Suppose the convergence is not uniform on  $I$ . Then there is  $\epsilon > 0$ , a subsequence  $f_{n_k}(x)$ , and a sequence  $x_k \in I$ , such that  $|f_{n_k}(x_k) - f(x_k)| \geq \epsilon$ . In the bounded and closed interval,  $x_k$  has a convergent subsequence. Without loss of generality, we may assume that  $x_k$  already converges to  $c \in I$ . For fixed  $n$ , we have  $n_k > n$  for sufficiently big  $k$ . Then by the monotone assumption, we have

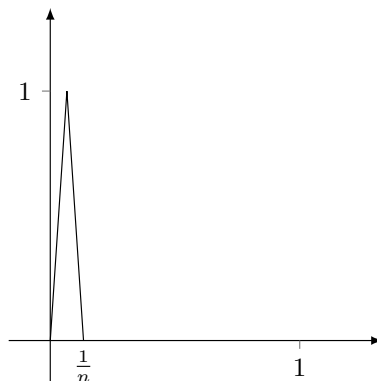
$$n_k > n \implies |f_n(x_k) - f(x_k)| \geq |f_{n_k}(x_k) - f(x_k)| \geq \epsilon.$$

Since the  $n_k > n$  for sufficiently big  $k$ , by the continuity of  $f_n$  and  $f$  at  $c$ , we may take  $k \rightarrow \infty$  and get  $|f_n(c) - f(c)| \geq \epsilon$ . Since this contradicts  $\lim_{n \rightarrow \infty} f_n(c) = f(c)$ , we conclude that that  $f_n$  converges to  $f$  uniformly on  $I$ .  $\square$

**Example 5.4.4.** Since  $x^n$  are continuous and the limit function in Example 5.3.6 is not continuous at  $1^-$ , we may use Proposition 5.4.2 to conclude that the convergence cannot be uniform on  $(r, 1)$  for any  $0 < r < 1$ .

We also note that  $x^n$  is decreasing in  $n$  for any  $x \in [0, r]$ , and the limit function 0 on the interval is continuous. By Dini's Theorem, the convergence is uniform on  $[0, r]$ .

**Example 5.4.5.** The function  $f_n(x)$  in Figure 5.4.1 is continuous. Although  $\lim_{n \rightarrow \infty} f_n(x) = 0$  is also continuous, the convergence is not uniform. Note that Dini's Theorem cannot be applied because  $f_n(x)$  is not monotone in  $n$ .



**Figure 5.4.1.** *A non-uniformly convergent sequence.*

**Example 5.4.6.** Applied to series of functions, Proposition 5.4.2 says that the sum of uniformly convergent series of continuous functions is continuous. Moreover, Proposition 5.4.3 says that, if the sum of non-negative continuous functions is continuous, then the convergence is uniform on bounded and closed interval.

The Riemann zeta function

$$\zeta(x) = 1 + \frac{1}{2^x} + \frac{1}{3^x} + \cdots + \frac{1}{n^x} + \cdots$$

is defined on  $(1, +\infty)$ . For any  $r > 1$ , we have  $0 < \frac{1}{n^x} \leq \frac{1}{n^r}$  on  $[r, +\infty)$ . By the comparison test, therefore,  $\sum \frac{1}{n^x}$  uniformly converges on  $[r, +\infty)$ . This implies that  $\zeta(x)$  is continuous on  $[r, +\infty)$  for any  $r > 1$ . Since  $r > 1$  is arbitrary, we conclude that  $\zeta(x)$  is continuous on  $(1, +\infty)$ .

**Example 5.4.7.** Let  $r_n$  be any sequence of distinct numbers. Let

$$u_n(x) = \begin{cases} 0, & \text{if } x < r_n, \\ 1, & \text{if } x \geq r_n. \end{cases}$$

Then by the comparison test, the series

$$\sum \frac{1}{2^n} u_n(x) = f(x) = \sum_{r_n \leq x} \frac{1}{2^n}$$

uniformly converges. If  $a$  is not any  $r_i$ , then all  $u_n(x)$  are continuous at  $a$ . By Proposition 5.4.2, this implies that  $f(x)$  is continuous at  $a$ . If  $a = r_j$ , then the same argument works for the series  $\sum_{n \neq j} \frac{1}{2^n} u_n(x)$ , so that the series is still continuous at  $a$ . However, since  $u_j(x)$  is not continuous at  $a$ , we know  $f(x) = \frac{1}{2^j} u_j(x) + \sum_{n \neq j} \frac{1}{2^n} u_n(x)$  is not continuous at  $a$ .

We conclude that  $f(x)$  is an increasing function that is continuous away from  $r_i$  and not continuous at all  $r_i$ . If we take the sequence to be all rational numbers, then  $f(x)$  is a strictly increasing function that is continuous at irrational numbers and not continuous at rational numbers.

Exercise 5.72. Is there a one point version of Dini's Theorem?

Exercise 5.73. State and prove the two variable function version of Propositions 5.4.2 and 5.4.3.

Exercise 5.74. Explain that the function  $f(x)$  in Example 5.4.7 is right continuous.

Exercise 5.75. Use Dini's Theorem to explain that the sequence  $\left(1 + \frac{x}{n}\right)^n$  uniformly converges on  $[-R, R]$  for any  $R > 0$ .

Exercise 5.76. A subset  $Y$  of  $X$  is dense if every element of  $X$  is the limit of a sequence in  $Y$ . Prove that if a sequence of continuous functions  $f_n(x)$  on  $X$  converges uniformly on  $Y$ , then it converges uniformly on  $X$ . In particular, the limit function is continuous on  $X$ .

Exercise 5.77. Suppose  $f_n$  uniformly converges to  $f$ . Prove that if  $f_n$  are uniformly continuous, then  $f$  is uniformly continuous.

## Uniform Convergence and Integration

**Proposition 5.4.4.** *Suppose  $f_n$  are integrable on a bounded interval  $[a, b]$  and uniformly converges to  $f$ . Then  $f$  is integrable and*

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \int_a^b f_n(x)dx.$$

The conclusion is the exchange of integration and limit

$$\int_a^b \lim_{n \rightarrow \infty} f_n(x)dx = \lim_{n \rightarrow \infty} \int_a^b f_n(x)dx.$$

Since the integration is defined as certain limit, the property is again the exchange of two limits. Specifically, the Riemann sums  $S(P, f_n) = \sigma_n(P)$  and  $S(P, f) = \sigma(P)$  may be considered as functions of the "variable"  $P$  (in fact the variable  $(P, x_i^*)$ , including the sample points), and the equality we wish to achieve is

$$\lim_{\|P\| \rightarrow 0} \lim_{n \rightarrow \infty} \sigma_n(P) = \lim_{n \rightarrow \infty} \lim_{\|P\| \rightarrow 0} \sigma_n(P).$$

For fixed  $P$  (including fixed sample points), we have

$$|f_n - f| < \epsilon \text{ on } [a, b] \implies |\sigma_n(P) - \sigma(P)| = |S(P, f_n - f)| < \epsilon(b - a).$$

Therefore the uniform convergence of  $f_n$  implies the uniform convergence of  $\sigma_n(P)$ , and we may apply Theorem 5.4.1 to get the equality of two repeated limits.

**Example 5.4.8.** The sequence  $f_n$  in Figure 5.4.1 converges to  $f = 0$  but not uniformly. Yet the conclusion of Proposition 5.4.4 still holds

$$\lim_{n \rightarrow \infty} \int_0^1 f_n(x)dx = \lim_{n \rightarrow \infty} \frac{1}{2n} = 0, \quad \int_0^1 \lim_{n \rightarrow \infty} f_n(x)dx = \int_0^1 0dx = 0.$$

On the other hand, the conclusion fails for  $\lim_{n \rightarrow \infty} n f_n = 0$

$$\lim_{n \rightarrow \infty} \int_0^1 n f_n(x) dx = \lim_{n \rightarrow \infty} \frac{1}{2} = \frac{1}{2}, \quad \int_0^1 \lim_{n \rightarrow \infty} n f_n(x) dx = \int_0^1 0 dx = 0.$$

The difference is that  $f_n$  is uniformly bounded, while  $n f_n$  is not uniformly bounded.

**Example 5.4.9.** The sequence  $x^n$  in Example 5.3.6 does not converge uniformly on  $[0, 1]$ . However, we still have

$$\lim_{n \rightarrow \infty} \int_0^1 x^n dx = \lim_{n \rightarrow \infty} \frac{1}{n+1} = 0 = \int_0^1 g(x) dx.$$

In fact, the Dominant Convergence Theorem (Theorem 10.4.4) in Lebesgue integration theory tells us that the equality  $\lim_{n \rightarrow \infty} \int_a^b f_n dx = \int_a^b f dx$  always holds as long as  $f_n$  are uniformly bounded, and the right side makes sense (i.e.,  $f$  is integrable).

**Example 5.4.10.** Let  $r_n, n = 1, 2, \dots$ , be all the rational numbers. Then the functions

$$f_n(x) = \begin{cases} 1, & \text{if } x = r_1, r_2, \dots, r_n, \\ 0, & \text{otherwise,} \end{cases}$$

are integrable. However,  $\lim_{n \rightarrow \infty} f_n(x) = D(x)$  is the Dirichlet function, which is not Riemann integrable. Of course the convergence is not uniform.

The example shows that the limit of Riemann integrable functions may not be Riemann integrable. The annoyance will be resolved by the introduction of Lebesgue integral, which extends the Riemann integral and allows more functions to be integrable. Then the Dirichlet function will be Lebesgue integrable with integral value 0, and the integration and the limit still exchange.

**Exercise 5.78.** Prove Proposition 5.4.4 by using the idea after the proposition and the proof of Theorem 5.4.1.

**Exercise 5.79.** State and prove the two variable function version of Proposition 5.4.4.

**Exercise 5.80.** Suppose  $f_n$  is integrable on a bounded interval  $[a, b]$  and  $\lim_{n \rightarrow \infty} f_n = f$  uniformly. Prove that the convergence of  $\lim_{n \rightarrow \infty} \int_a^x f_n(t) dt = \int_a^x f(t) dt$  is uniform for  $x \in [a, b]$ .

**Exercise 5.81.** Extend Proposition 5.4.4 to the Riemann-Stieltjes integral.

**Exercise 5.82.** Suppose  $f$  is Riemann-Stieltjes integrable with respect to each  $\alpha_n$ . Will the uniform convergence of  $\alpha_n$  tell you something about the limit of  $\int_a^b f d\alpha_n$ ?

## Uniform Convergence and Differentiation

**Proposition 5.4.5.** Suppose  $f_n$  are differentiable on a bounded interval, such that  $f_n(x_0)$  converges at some point  $x_0$  and  $f'_n$  uniformly converges to  $g$ . Then  $f_n$  uni-

formly converges and

$$\left( \lim_{n \rightarrow \infty} f_n \right)' = g.$$

The conclusion is the exchange of derivative and limit

$$\left( \lim_{n \rightarrow \infty} f_n \right)' = \lim_{n \rightarrow \infty} f'_n.$$

At  $x_0$ , this means the exchange of two limits

$$\lim_{x \rightarrow x_0} \lim_{n \rightarrow \infty} \frac{f_n(x) - f_n(x_0)}{x - x_0} = \lim_{n \rightarrow \infty} \lim_{x \rightarrow x_0} \frac{f_n(x) - f_n(x_0)}{x - x_0}.$$

The interpretation inspires the following proof.

*Proof.* Let  $g_n(x) = \frac{f_n(x) - f_n(x_0)}{x - x_0}$  for  $x \neq x_0$  and  $g_n(x_0) = f'_n(x_0)$ . We apply the Mean Value Theorem to  $f_m(x) - f_n(x)$  and get

$$g_m(x) - g_n(x) = \frac{(f_m(x) - f_n(x)) - (f_m(x_0) - f_n(x_0))}{x - x_0} = f'_m(c) - f'_n(c)$$

for some  $c$  between  $x_0$  and  $x$ . The equality also holds for  $x = x_0$ . Then the Cauchy criterion for the uniform convergence of  $f'_n(x)$  implies the Cauchy criterion for the uniform convergence of  $g_n(x)$ , so that  $g_n(x)$  uniformly converges. Then by the convergence of  $f_n(x_0)$ , the sequence  $f_n(x) = f_n(x_0) + (x - x_0)g_n(x)$  also uniformly converges on the bounded interval. Moreover, by Theorem 5.4.1,  $\lim_{x \rightarrow x_0} g(x)$  converges and

$$\begin{aligned} f'(x_0) &= \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \rightarrow x_0} \lim_{n \rightarrow \infty} \frac{f_n(x) - f_n(x_0)}{x - x_0} \\ &= \lim_{x \rightarrow x_0} \lim_{n \rightarrow \infty} g_n(x) = \lim_{n \rightarrow \infty} \lim_{x \rightarrow x_0} g_n(x) = \lim_{n \rightarrow \infty} f'_n(x_0). \end{aligned}$$

So the equality for the derivative is proved at  $x_0$ . Since the first consequence tells us that  $f_n(x)$  converges everywhere, so we can actually take  $x_0$  to be anywhere in the interval. Therefore the equality for the derivative holds anywhere in the interval.  $\square$

**Example 5.4.11.** In Example 5.3.8, we know that  $\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$  uniformly on  $[-R, R]$  for any  $R > 0$ . The derivative sequence is

$$\frac{d}{dx} \left(1 + \frac{x}{n}\right)^n = n \left(1 + \frac{x}{n}\right)^{n-1} \frac{1}{n} = \frac{n}{n+x} \left(1 + \frac{x}{n}\right)^n.$$

Since both  $\frac{n}{n+x}$  and  $\left(1 + \frac{x}{n}\right)^n$  are bounded and uniformly converge on  $[-R, R]$ , we get

$$(e^x)' = \lim_{n \rightarrow \infty} \frac{d}{dx} \left(1 + \frac{x}{n}\right)^n = e^x \text{ on } (-R, R).$$

Since  $R$  is arbitrary, we get  $(e^x)' = e^x$  for all  $x$ .

**Example 5.4.12.** Applied to series, Proposition 5.4.5 says that, if the derivative series uniformly converges, then we can take term by term derivative of the sum of series.

The terms by term derivation of the Riemann zeta function in Example 5.4.6 is

$$-\sum_{n=2}^{\infty} \frac{\log n}{n^x} = -\frac{\log 2}{2^x} - \frac{\log 3}{3^x} + \dots - \frac{\log n}{n^x} - \dots$$

For any  $r > 1$ , choose  $r'$  satisfying  $r > r' > 1$ . Then  $0 < \frac{\log n}{n^x} \leq \frac{\log n}{n^r} < \frac{1}{n^{r'}}$  for  $x \geq r$  and sufficiently big  $n$ . By the comparison test,  $\sum \frac{\log n}{n^x}$  uniformly converges on  $[r, +\infty)$  for any  $r > 1$ . By Proposition 5.4.5, this implies that  $\zeta(x)$  is differentiable and  $\zeta'(x) = -\sum \frac{\log n}{n^x}$  for  $x > 1$ . Further argument shows that  $\zeta(x)$  has derivative of any order.

**Example 5.4.13** (Continuous and Nowhere Differentiable Function<sup>29</sup>). Let  $h(x)$  be given by  $h(x) = |x|$  on  $[-1, 1]$  and  $h(x+2) = h(x)$  for any  $x$ . The function is continuous and satisfies  $0 \leq h(x) \leq 1$ . By the comparison test, therefore, the series

$$f(x) = \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n h(4^n x)$$

uniformly converges and the sum  $f(x)$  is continuous. However, we will show that  $f(x)$  is not differentiable anywhere.

Let  $\delta_k = \pm \frac{1}{2 \cdot 4^k}$ . For any  $n$ , by  $|h(x) - h(y)| \leq |x - y|$ , we have

$$\left| \frac{h(4^n(a + \delta_k)) - h(4^n a)}{\delta_k} \right| \leq \frac{4^n \delta_k}{\delta_k} = 4^n.$$

For  $n > k$ ,  $4^n \delta_k$  is a multiple of 2, and we have

$$\frac{h(4^n(a + \delta_k)) - h(4^n a)}{\delta_k} = 0.$$

For  $n = k$ , we have  $4^k \delta_k = \pm \frac{1}{2}$ . By choosing  $\pm$  sign so that there is no integer between  $4^k a$  and  $4^k a \pm \frac{1}{2}$ , we can make sure that  $|h(4^k(a + \delta_k)) - h(4^k a)| = |4^k(a + \delta_k) - 4^k a| = \frac{1}{2}$ . Then

$$\left| \frac{h(4^k(a + \delta_k)) - h(4^k a)}{\delta_k} \right| = 4^k.$$

Thus for any fixed  $a$ , by choosing a sequence  $\delta_k$  with suitable  $\pm$  sign, we get

$$\left| \frac{f(a + \delta_k) - f(a)}{\delta_k} \right| \geq \left(\frac{3}{4}\right)^k 4^k - \sum_{n=0}^{k-1} \left(\frac{3}{4}\right)^n 4^n = 3^k - \frac{3^k - 1}{3 - 1} = \frac{3^k + 1}{2}.$$

This implies that  $\lim_{\delta \rightarrow 0} \frac{f(a + \delta) - f(a)}{\delta}$  diverges.

<sup>29</sup>For a generalisation of the example, see “Constructing Nowhere Differentiable Functions from Convex Functions” by Cater, Real Anal. Exchange **28** (2002/03) 617-621.

Exercise 5.83. State and prove the two variable function version of Proposition 5.4.5.

Exercise 5.84. Justify the equalities.

1.  $\int_0^1 x^{ax} dx = \sum_{n=1}^{\infty} \frac{(-a)^{n-1}}{n^n}.$
2.  $\int_{\lambda a}^a \left( \sum_{n=0}^{\infty} \lambda^n \tan \lambda^n x \right) dx = -\log |\cos a|$  for  $|a| < \frac{\pi}{2}, |\lambda| < 1.$
3.  $\int_x^{+\infty} (\zeta(t) - 1) dt = \sum_{n=2}^{\infty} \frac{1}{n^x \log n}$  for  $x > 1.$

Exercise 5.85. Find the places where the series converge and has derivatives. Also find the highest order of the derivative.

1.  $\sum \frac{1}{n(\log n)^x}.$
2.  $\sum \left( x + \frac{1}{n} \right)^n.$
3.  $\sum_{n=-\infty}^{+\infty} \frac{1}{|n-x|^p}.$
4.  $\sum \frac{(-1)^n}{n^x}.$

Exercise 5.86. Find  $\lim_{x \rightarrow 0} \frac{1}{x} \left( \sum_{n=1}^{\infty} (-1)^n \frac{1}{n+x^2} \right)'.$

Exercise 5.87. In case  $x_0$  is the right end of the bounded interval in Proposition 5.4.5, prove that the conclusion holds for the left derivative at  $x_0$ .

Exercise 5.88. Suppose  $\sum \frac{1}{a_n}$  absolutely converges. Prove that  $\sum \frac{1}{x - a_n}$  converges to a function that has derivatives of any order away from all  $a_n$ .

## Power Series

A *power series* is a series of the form

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots,$$

or of the more general form  $\sum_{n=0}^{\infty} a_n (x - x_0)^n$ . Taylor series are power series.

**Theorem 5.4.6.** *Let*

$$R = \frac{1}{\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}}.$$

1. *The power series  $\sum_{n=0}^{\infty} a_n x^n$  absolutely converges for  $|x| < R$  and diverges for  $|x| > R$ .*
2. *If  $\sum_{n=0}^{\infty} a_n r^n$  converges, then the power series uniformly converges on  $[0, r]$ .*

The number  $R$  is called the *radius of convergence* for the power series.

In the second statement, we allow  $r$  to be negative, and  $[0, r]$  is really  $[r, 0]$  in this case. A consequence of the theorem is that the power series uniformly converges on  $[-r, r]$  for any  $0 < r < R$ . Moreover, if  $\sum_{n=0}^{\infty} a_n R^n$  converges, then the power series uniformly converges on  $[0, R]$ . If  $\sum_{n=0}^{\infty} a_n (-R)^n$  converges, then the power series uniformly converges on  $[-R, 0]$ . The uniform convergence up to  $\pm R$  is called Abel's Theorem.

*Proof.* If  $|x| < R$ , then  $\lim_{n \rightarrow \infty} |x| \sqrt[n]{|a_n|} < 1$ . Fix  $r$  satisfying  $\lim_{n \rightarrow \infty} |x| \sqrt[n]{|a_n|} < r < 1$ . By Proposition 1.5.4, there are only finitely many  $n$  satisfying  $|x| \sqrt[n]{|a_n|} > r$ . This means that  $|x| \sqrt[n]{|a_n|} \leq r$  for sufficiently big  $n$ , or  $|a_n x^n| \leq r^n$  for sufficiently big  $n$ . By the convergence of  $\sum r^n$  and the comparison test, we find that  $\sum a_n x^n$  absolutely converges.

If  $|x| > R$ , then similar argument using Proposition 1.5.4 shows that there are infinitely many  $n$  satisfying  $|a_n x^n| > 1$ . This implies  $a_n x^n$  does not converge to 0, and therefore  $\sum a_n x^n$  diverges.

Suppose  $\sum a_n r^n$  converges. Then the Abel test in Proposition 5.3.4 may be applied to  $u_n(x) = \frac{x^n}{r^n}$  and  $v_n(x) = a_n r^n$  on  $[0, r]$ . The functions  $u_n(x)$  are all bounded by 1, and the sequence  $u_n(x)$  is decreasing for each  $x \in [0, r]$ . The convergence of  $\sum v_n(x)$  is uniform because the series is independent of  $x$ . The conclusion is the uniform convergence of  $\sum u_n(x)v_n(x) = \sum a_n x^n$  on  $[0, r]$ .  $\square$

The domain of the sum function  $f(x) = \sum a_n x^n$  is an interval with left end  $-R$  and right end  $R$ . Combining Proposition 5.4.2 with the second part of Theorem 5.4.6, we find that  $f(x)$  is continuous wherever it is defined.

Similarly, we may combine Propositions 5.4.4 and 5.4.5 with Theorem 5.4.6 to find that the integration and differentiation of a power series can be carried out term by term

$$(a_0 + a_1 x + \cdots + a_n x^n + \cdots)' = a_1 + 2a_2 x + \cdots + n a_n x^{n-1} + \cdots,$$

$$\int_0^x (a_0 + a_1 t + \cdots + a_n t^n + \cdots) dt = a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3 + \cdots + \frac{a_n}{n+1} x^{n+1} + \cdots.$$

The two series also have  $R$  as the radius of convergence, and the equalities hold wherever the power series on the right converges.

**Example 5.4.14.** By the equality

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \cdots \text{ on } (-1, 1),$$

we get

$$\frac{1}{1+x} = 1 - x + x^2 - \cdots + (-1)^n x^n + \cdots \text{ on } (-1, 1).$$

By integration, we get

$$\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots + \frac{(-1)^n}{n+1}x^{n+1} + \cdots \text{ on } (-1, 1).$$



Since the series on the right also converges at  $x = 1$ , the right side is continuous at  $1^-$ , and

$$\begin{aligned} 1 - \frac{1}{2} + \frac{1}{3} + \cdots + \frac{(-1)^{n+1}}{n} + \cdots &= \lim_{x \rightarrow 1^-} \left( x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots + \frac{(-1)^n}{n+1}x^{n+1} + \cdots \right) \\ &= \lim_{x \rightarrow 1^-} \log(1+x) \quad (\text{sum of right for } |x| < 1) \\ &= \log 2. \quad (\text{continuity of } \log(1+x)) \end{aligned}$$

Therefore the Taylor expansion equality for  $\log(1+x)$  extends to  $x = 1$ .

**Exercise 5.89.** Prove that the radius of convergence satisfies

$$\liminf_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right| \leq R \leq \limsup_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right|.$$

Then show that the radius of convergence for the Taylor series of  $(1+x)^p$  ( $p$  is not a natural number) and  $\log(1+x)$  is 1.

**Exercise 5.90.** Prove that the power series obtained from term by term differentiation or integration has the same radius of convergence. Can you make a direct argument without using the formula for the radius?

**Exercise 5.91.** Prove that if the radius of convergence for the power series  $\sum a_n x^n$  is nonzero, then the sum function  $f(x)$  satisfies  $f^{(n)}(0) = n!a_n$ .

**Exercise 5.92.** Use the Taylor series of  $\arctan x$  to show that

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots = \frac{\pi}{4}.$$

Then show

$$1 + \frac{1}{2} - \frac{1}{3} - \frac{1}{4} + \frac{1}{5} + \frac{1}{6} - \frac{1}{7} - \frac{1}{8} + \cdots = \frac{\pi}{4} + \frac{\log 2}{2}.$$

**Exercise 5.93.** Discuss the convergence of the Taylor series of  $\arcsin x$  at the radius of convergence.

**Exercise 5.94.** The series  $\sum z_n$  with  $z_n = x_0 y_n + x_1 y_{n-1} + \cdots + x_n y_0$  is obtained by combining terms in the diagonal arrangement of the product of the series  $\sum x_n$  and  $\sum y_n$ . By considering the power series  $\sum x_n t^n$ ,  $\sum y_n t^n$ ,  $\sum z_n t^n$  at  $t = 1$ , prove that if  $\sum x_n$ ,  $\sum y_n$  and  $\sum z_n$  converge, then  $\sum z_n = (\sum x_n)(\sum y_n)$ . Compare with Exercise 5.49.

## 5.5 Additional Exercise

### Approximate Partial Sum by Integral

The proof of Proposition 5.2.2 gives an estimation of the partial sum for  $\sum f(n)$  by the integral of  $f(x)$  on suitable intervals. The idea leads to an estimation of  $n!$  in Exercise 5.33. In what follows, we study the approximation in general.

Suppose  $f(x)$  is a decreasing function on  $[1, +\infty)$  satisfying  $\lim_{x \rightarrow +\infty} f(x) = 0$ . Denote

$$d_n = f(1) + f(2) + \cdots + f(n) - \int_1^n f(x) dx.$$

By Exercise 5.32,  $d_n$  is decreasing and converges to a limit  $\gamma \in [0, f(1)]$ .

**Exercise 5.95.** Prove that if  $f(x)$  is convex, then  $d_n \geq \frac{1}{2}(f(1) + f(n))$ . This implies  $\gamma \geq \frac{1}{2}f(1)$ .

**Exercise 5.96.** By using Exercise 4.103, prove that if  $f(x)$  is convex and differentiable, then for  $m > n$ , we have

$$\left| d_n - d_m - \frac{1}{2}(f(n) - f(m)) \right| \leq \frac{1}{8}(f'(n) - f'(m)).$$

**Exercise 5.97.** Prove that if  $f(x)$  is convex and differentiable, then

$$\left| d_n - \gamma - \frac{1}{2}f(n) \right| \leq -\frac{1}{8}f'(n).$$

**Exercise 5.98.** By using Exercise 4.106, prove that if  $f(x)$  has second order derivative, then

$$\frac{1}{24} \sum_{k=m}^{n-1} \inf_{[k, k+1]} f'' \leq -d_n + d_m + \frac{1}{2}(f(n) - f(m)) + \frac{1}{8}(f'(n) - f'(m)) \leq \frac{1}{24} \sum_{k=m}^{n-1} \sup_{[k, k+1]} f''.$$

**Exercise 5.99.** Let  $\gamma$  be the Euler-Mascheroni constant in Exercise 5.32. Prove that

$$\frac{1}{24(n+1)^2} \leq 1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n - \gamma - \frac{1}{2n} + \frac{1}{8n^2} \leq \frac{1}{24(n-1)^2}.$$

This implies

$$1 + \frac{1}{2} + \cdots + \frac{1}{n} = \log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + o\left(\frac{1}{n^2}\right).$$

**Exercise 5.100.** Estimate  $1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}}$ .

### Bertrand<sup>30</sup> Test

The Bertrand test is obtained by applying the general ratio test in Exercise 5.26 to  $y_n = \frac{1}{n(\log n)^p}$ .

**Exercise 5.101.** Prove that  $\sum \frac{1}{n(\log n)^p}$  converges if and only if  $p > 1$ .

**Exercise 5.102.** Prove that if  $\left| \frac{x_{n+1}}{x_n} \right| \leq 1 - \frac{1}{n} - \frac{p}{n \log n}$  for some  $p > 1$  and sufficiently big  $n$ , then  $\sum x_n$  absolutely converges.

<sup>30</sup>Joseph Louis François Bertrand, born 1822 and died 1900 in Paris (France).

Exercise 5.103. Prove that if  $\left| \frac{x_{n+1}}{x_n} \right| \geq 1 - \frac{1}{n} - \frac{p}{n \log n}$  for some  $p < 1$  and sufficiently big  $n$ , then  $\sum |x_n|$  diverges.

Exercise 5.104. Find the limit version of the test.

Exercise 5.105. What can you say in case  $\left| \frac{x_{n+1}}{x_n} \right| = 1 - \frac{1}{n} - \frac{1}{n \log n} + o\left(\frac{1}{n \log n}\right)$ ?

Exercise 5.106. Rephrase the test in terms of the quotient  $\left| \frac{x_n}{x_{n+1}} \right|$ .

### Kummer<sup>31</sup> Test

Exercise 5.107. Prove that if there are  $c_n > 0$  and  $\delta > 0$ , such that  $c_n - c_{n+1} \left| \frac{x_{n+1}}{x_n} \right| \geq \delta$  for sufficiently big  $n$ , then  $\sum x_n$  absolutely converges.

Exercise 5.108. Prove that if  $x_n > 0$ , and there are  $c_n > 0$ , such that  $c_n - c_{n+1} \frac{x_{n+1}}{x_n} \leq 0$  and  $\sum \frac{1}{c_n}$  diverges, then  $\sum x_n$  diverges.

Exercise 5.109. Prove that if  $\sum x_n$  absolutely converges, then there are  $c_n > 0$ , such that  $c_n - c_{n+1} \left| \frac{x_{n+1}}{x_n} \right| \geq 1$  for all  $n$ .

Exercise 5.110. Derive ratio, Raabe and Bertrand tests from the Kummer test.

### Infinite Product

An *infinite product* is

$$\prod_{n=1}^{\infty} x_n = x_1 x_2 \cdots x_n \cdots$$

The infinite product converges if the partial *partial product*  $p_n = x_1 x_2 \cdots x_n$  converges to some  $p \neq 0$ , and we denote  $p = \prod x_n$ .

Exercise 5.111. Explain that the assumption  $p \neq 0$  implies that a convergent infinite product satisfies  $\lim_{n \rightarrow \infty} x_n = 1$ .

Exercise 5.112. Explain that the convergence is not affected by dropping finitely many terms. This means that, as far as the convergence is concerned, we may assume all  $x_n > 0$  (this will be assumed in all subsequent exercises). Prove that, under the assumption of all  $x_n > 0$ , an infinite product converges if and only if  $\sum \log x_n$  converges.

Exercise 5.113. Compute the convergent infinite products.

<sup>31</sup>Ernst Eduard Kummer, born 1810 in Sorau (Prussia, now Poland), died 1893 in Berlin (Germany). Besides numerous contributions to analysis, Kummer made fundamental contributions to modern algebra.

1.  $\prod_{n=1}^{\infty} \left(1 + \frac{1}{n}\right)$ .
2.  $\prod_{n=2}^{\infty} \left(1 + \frac{(-1)^n}{n}\right)$ .
3.  $\prod_{n=2}^{\infty} \frac{n-1}{n+1}$ .
4.  $\prod_{n=2}^{\infty} \frac{n^2 - n + 1}{n^2 + n + 1}$ .
5.  $\prod_{n=2}^{\infty} \frac{n^3 - 1}{n^3 + 1}$ .
6.  $\prod_{n=1}^{\infty} 2^{\frac{1}{n}}$ .
7.  $\prod_{n=0}^{\infty} 2^{3^{-n}}$ .
8.  $\prod_{n=0}^{\infty} 2^{\frac{(-1)^n}{n!}}$ .
9.  $\prod_{n=1}^{\infty} \cos \frac{x}{2^n}$ .

Exercise 5.114. Establish properties for infinite products similar to Exercises 5.13 and 5.14.

Exercise 5.115. State the Cauchy criterion for the convergence of infinite product.

Exercise 5.116. Suppose  $x_n \neq -1$ .

1. Prove that if  $\sum x_n$  converges, then  $\prod(1+x_n)$  converges if and only if  $\sum x_n^2$  converges.
2. Prove that if  $\sum x_n^2$  converges, then  $\prod(1+x_n)$  converges if and only if  $\sum x_n$  converges.

Exercise 5.117. Construct a convergent series  $\sum x_n$ , such that the series  $\sum x_n^2$  diverges. By Exercise 5.116, we see that the convergence of  $\sum x_n$  does not necessarily imply the convergence of  $\prod(1+x_n)$ .

Exercise 5.118. Suppose  $x_n$  decreases and  $\lim_{n \rightarrow \infty} x_n = 1$ . Prove that the “alternating infinite product” (like Leibniz test in Exercise 5.37)  $\prod x_n^{(-1)^n}$  converges. Then find suitable  $x_n$ , such that the series  $\sum y_n$  defined by  $1 + y_n = x_n^{(-1)^n}$  diverges. This shows that the convergence of  $\prod(1+y_n)$  does not necessarily imply the convergence of  $\sum y_n$ .

### Absolute Convergence of Infinite Product

Suppose  $x_n \neq -1$ . An infinite product  $\prod(1+x_n)$  *absolutely converges* if  $\prod(1+|x_n|)$  converges.

Exercise 5.119. Prove that  $\prod(1+x_n)$  absolutely converges if and only if the series  $\sum x_n$  absolutely converges. Moreover,  $\prod(1+|x_n|) = +\infty$  if and only if  $\sum x_n = +\infty$ .

Exercise 5.120. Prove that if the infinite product absolutely converges, then the infinite product converges.

Exercise 5.121. Suppose  $0 < x_n < 1$ . Prove that  $\prod(1+x_n)$  converges if and only if  $\prod(1-x_n)$  converges. Moreover,  $\prod(1+x_n) = +\infty$  if and only if  $\prod(1-x_n) = 0$ .

Exercise 5.122. State the analogue of Theorem 5.2.6 for infinite product.

### Ratio Rule

By relating  $x_n$  to the partial product of  $\prod \frac{x_{n+1}}{x_n}$ , the limit of the sequence  $x_n$  can be studied by considering the ratio  $\frac{x_{n+1}}{x_n}$ . This leads to the extension of the ratio rules in Exercises 1.58, 3.142, 3.145.

Exercise 5.123. Suppose  $\left| \frac{x_{n+1}}{x_n} \right| \leq 1 - y_n$  and  $0 < y_n < 1$ . Use Exercises 5.119 and 5.121 to prove that if  $\sum y_n = +\infty$ , then  $\lim_{n \rightarrow \infty} x_n = 0$ .

Exercise 5.124. Suppose  $\left| \frac{x_{n+1}}{x_n} \right| \geq 1 + y_n$  and  $0 < y_n < 1$ . Prove that if  $\sum y_n = +\infty$ , then  $\lim_{n \rightarrow \infty} x_n = \infty$ .

Exercise 5.125. Suppose  $1 - y_n \leq \frac{x_{n+1}}{x_n} \leq 1 + z_n$  and  $0 < y_n, z_n < 1$ . Prove that if  $\sum y_n$  and  $\sum z_n$  converge, then  $\lim_{n \rightarrow \infty} x_n$  converges to a nonzero limit.

Exercise 5.126. Study  $\lim_{n \rightarrow \infty} \frac{(n+a)^{n+\frac{1}{2}}}{(\pm e)^n n!}$ , the case not yet settled in Exercise 3.143.

### Riemann Zeta Function and Number Theory

The Riemann zeta function  $\zeta(x) = \sum_{n=0}^{\infty} \frac{1}{n^x}$  is introduced in Example 5.4.6. The function is a fundamental tool for number theory.

Let

$$p_n: 2, 3, 5, 7, 11, 13, 17, 19, \dots$$

be all prime numbers in increasing order. Let  $S_n$  be the set of natural numbers whose prime factors are among  $p_1, p_2, \dots, p_n$ . For example,  $20 \notin S_2$  and  $20 \in S_3$  because the prime factors of 20 are 2 and 5.

Exercise 5.127. Prove that  $\prod_{i=1}^k \left(1 - \frac{1}{p_i^x}\right)^{-1} = \sum_{n \in S_k} \frac{1}{n^x}$ .

Exercise 5.128. Prove that the infinite product  $\prod \left(1 - \frac{1}{p_n^x}\right)$  converges if and only if the series  $\sum \frac{1}{n^x}$  converges.

Exercise 5.129. With the help of Exercises 5.119 and 5.121, prove that  $\sum \frac{1}{p_n^x}$  converges if and only if  $x > 1$ .

Exercise 5.130. Exercise 5.129 tells us that  $\sum \frac{1}{p_n}$  diverges. By expressing numbers in base  $10^6$  and using the idea of Exercise 5.31, prove that there must be a prime number, such that its decimal expression contains the string 123456. Can you extend the conclusion to a general result?

**The Series**  $\sum_{n=1}^{\infty} \frac{a_n}{n^x}$

The series has many properties similar to the power series.

Exercise 5.131. Use the Abel test in Proposition 5.3.4 to show that if  $\sum \frac{a_n}{n^r}$  converges,

then  $\sum \frac{a_n}{n^x}$  uniformly converges on  $[r, +\infty)$ , and  $\sum \frac{a_n}{n^r} = \lim_{x \rightarrow r^+} \sum \frac{a_n}{n^x}$ .

**Exercise 5.132.** Prove that there is  $R$ , such that  $\sum \frac{a_n}{n^x}$  converges on  $(R, +\infty)$  and diverges on  $(-\infty, R)$ . Moreover, prove that  $R \geq \overline{\lim}_{n \rightarrow \infty} \frac{\log |a_n|}{\log n}$ .

**Exercise 5.133.** Prove that we can take term wise integration and derivative of any order of the series on  $(R, +\infty)$ .

**Exercise 5.134.** Prove that there is  $R'$ , such that the series absolutely converges on  $(R', +\infty)$  and absolutely diverges on  $(R', +\infty)$ . Moreover, prove that  $\overline{\lim}_{n \rightarrow \infty} \frac{\log |a_n|}{\log n} + 1 \geq R' \geq \underline{\lim}_{n \rightarrow \infty} \frac{\log |a_n|}{\log n} + 1$ .

**Exercise 5.135.** Give an example to show the inequalities in Exercises 5.132 and 5.134 can be strict.

### An Example by Borel

Suppose  $a_n > 0$  and  $\sum \sqrt{a_n}$  converges. Suppose the set  $\{r_n\}$  is all the rational numbers in  $[0, 1]$ . We study the convergence of the series  $\sum \frac{a_n}{|x - r_n|}$  for  $x \in [0, 1]$ ?

**Exercise 5.136.** Prove that if  $x \notin \cup_n (r_n - c\sqrt{a_n}, r_n + c\sqrt{a_n})$ , then the series converges.

**Exercise 5.137.** Use Heine-Borel theorem to prove that if  $\sum \sqrt{a_n} < \frac{1}{2c}$ , then  $[0, 1] \not\subset \cup_n (r_n - c\sqrt{a_n}, r_n + c\sqrt{a_n})$ . By Exercise 5.136, this implies that the series converges for some  $x \in [0, 1]$ .

### Continuous But Not Differentiable Function

**Exercise 5.138.** Let  $a_n$  be any sequence of points. Let  $|b| < 1$ . Prove that the function  $f(x) = \sum b^n |x - a_n|$  is continuous and is not differentiable precisely at  $a_n$ .

**Exercise 5.139 (Riemann).** Let  $((x))$  be the periodic function with period 1, determined by  $((x)) = x$  for  $-\frac{1}{2} < x < \frac{1}{2}$  and  $((\frac{1}{2})) = 0$ . Let  $f(x) = \sum_{n=1}^{\infty} \frac{((nx))}{n^2}$ .

1. Prove that  $f(x)$  is not continuous precisely at rational numbers with even denominators, i.e., number of the form  $r = \frac{a}{2b}$ , where  $a$  and  $b$  are odd integers.
2. Compute  $f(r^+) - f(r)$  and  $f(r^-) - f(r)$  at discontinuous points. You may need  $\sum \frac{1}{n^2} = \frac{\pi^2}{6}$  for the precise value.
3. Prove that  $f(x)$  is integrable, and  $F(x) = \int_0^x f(t)dt$  is not differentiable precisely at rational numbers with even denominators.

**Exercise 5.140 (Weierstrass).** Suppose  $0 < b < 1$  and  $a$  is an odd integer satisfying  $ab > 1 + \frac{3\pi}{2}$ . Prove that  $\sum_{n=0}^{\infty} b^n \cos(a^n \pi x)$  is continuous but nowhere differentiable.

**Exercise 5.141 (Weierstrass).** Suppose  $0 < b < 1$  and  $\{a_n\}$  is a bounded countable set of numbers. Let  $h(x) = x + \frac{x}{2} \sin \log |x|$  and  $h(0) = 0$ . Prove that  $\sum_{n=1}^{\infty} b^n h(x - a_n)$  is continuous, strictly increasing, and is not differentiable precisely at  $a_n$ .

### Double Series

A double series  $\sum_{m,n \geq 1} x_{m,n}$  may converge in several different ways.

First, the double series *converges to sum*  $s$  if for any  $\epsilon > 0$ , there is  $N$ , such that

$$m, n > N \implies \left| \sum_{i=1}^m \sum_{j=1}^n x_{i,j} - s \right| < \epsilon.$$

Second, the double series has *repeated sum*  $\sum_m \sum_n x_{m,n}$  if  $\sum_n x_{m,n}$  converges for each  $m$  and the series  $\sum_m (\sum_n x_{m,n})$  again converges. Of course, there is another repeated sum  $\sum_n \sum_m x_{m,n}$ .

Third, a one-to-one correspondence  $k \in \mathbb{N} \mapsto (m(k), n(k)) \in \mathbb{N}^2$  arranges the double series into a single series  $\sum_k x_{m(k), n(k)}$ , and we may consider the sum of the single series.

Fourth, for any finite subset  $A \subset \mathbb{N}^2$ , we may define the partial sum

$$s_A = \sum_{(m,n) \in A} x_{m,n}.$$

Then for any sequence  $A_k$  of finite subsets satisfying  $A_k \subset A_{k+1}$  and  $\cup A_k = \mathbb{N}^2$ , we say the double series converges to  $s$  with respect to the sequence  $A_k$  if  $\lim_{k \rightarrow \infty} s_{A_k} = s$ . For example, we have the *spherical sum* by considering  $A_k = \{(m,n) \in \mathbb{N}^2 : m^2 + n^2 \leq k^2\}$  and the *triangular sum* by considering  $A_k = \{(m,n) \in \mathbb{N}^2 : m + n \leq k\}$ .

Finally, the double series *absolutely converges* (see Exercise 5.143 for the reason for the terminology) to  $s$  if for any  $\epsilon > 0$ , there is  $N$ , such that  $|s_A - s| < \epsilon$  for any  $A$  containing all  $(m,n)$  satisfying  $m \leq N$  and  $n \leq N$ .

**Exercise 5.142.** State the Cauchy criterion for the convergence of  $\sum x_{m,n}$  in the first sense. State the corresponding comparison test.

**Exercise 5.143.** Prove that  $\sum x_{m,n}$  absolutely converges in the final sense if and only if  $\sum |x_{m,n}|$  converges in the first sense.

**Exercise 5.144.** Prove that a double series  $\sum x_{m,n}$  absolutely converges if and only if all the arrangement series  $\sum_k x_{m(k), n(k)}$  converge. Moreover, the arrangement series have the same sum.

**Exercise 5.145.** Prove that a double series  $\sum x_{m,n}$  absolutely converges if and only if the double series converges with respect to all the sequences  $A_k$ . Moreover, the sums with respect to all the sequences  $A_k$  are the same.

**Exercise 5.146.** Prove that if a double series  $\sum x_{m,n}$  absolutely converges, then the two repeated sums converge and have the same value.

Exercise 5.147. If a double series does not converge absolutely, what can happen to various sums?

Exercise 5.148. Study the convergence and values of the following double series.

1.  $\sum_{m,n \geq 1} a^{mn}$ .
2.  $\sum_{m,n \geq 1} \frac{1}{(m+n)^p}$ .
3.  $\sum_{m,n \geq 1} \frac{(-1)^{m+n}}{(m+n)^p}$ .
4.  $\sum_{m,n \geq 2} \frac{1}{n^m}$ .

### Gamma Function

The *Gamma function* is

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

Exercise 5.149. Prove that the function is defined and continuous for  $x > 0$ .

Exercise 5.150. Prove  $\lim_{x \rightarrow 0^+} \Gamma(x) = \lim_{x \rightarrow +\infty} \Gamma(x) = +\infty$ .

Exercise 5.151. Prove the other formulae for the Gamma function

$$\Gamma(x) = 2 \int_0^\infty t^{2x-1} e^{-t^2} dt = a^x \int_0^\infty t^{x-1} e^{-at} dt.$$

Exercise 5.152. Prove the equalities for the Gamma function

$$\Gamma(x+1) = x\Gamma(x), \quad \Gamma(n) = (n-1)!.$$

### Beta Function

The *Beta function* is

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt.$$

Exercise 5.153. Prove that the function is defined for  $x, y > 0$ .

Exercise 5.154. Use the change of variable  $t = \frac{1}{1+u}$  to prove the other formula for the Beta function

$$B(x, y) = \int_0^\infty \frac{t^{y-1} dt}{(1+t)^{x+y}} = \int_0^1 \frac{t^{x-1} + t^{y-1}}{(1+t)^{x+y}} dt.$$

Exercise 5.155. Prove

$$B(x, y) = 2 \int_0^{\frac{\pi}{2}} \cos^{2x-1} t \sin^{2y-1} t dt.$$

Exercise 5.156. Prove the equalities

$$B(x, y) = B(x, y), \quad B(x+1, y) = \frac{x}{x+y} B(x, y).$$



## Chapter 6

# Multivariable Function

## 6.1 Limit in Euclidean Space

The extension of analysis from single variable to multivariable means going from the real line  $\mathbb{R}$  to the Euclidean space  $\mathbb{R}^n$ . While many concepts and results may be extended, the Euclidean space is more complicated in various aspects. The first complication is many possible choices of distance in a Euclidean space, which we will show to be all equivalent as far as the mathematical analysis is concerned. The second complication is that it is not sufficient to just do analysis on the rectangles, the obvious generalization of the intervals on the real line. For example, to analyze the temperature around the globe, we need to deal with a function defined on the 2-dimensional sphere inside the 3-dimensional Euclidean space. Thus we need to set up proper topological concepts such as closed subset, compact subset and open subset, that extend closed interval, bounded closed interval and open interval. Then we may extend the discussion about the limit and continuity of single variable functions to multivariable functions defined on suitable subsets. The only properties that cannot be extended are those dealing with orders among real numbers, such as the monotone properties.

### Euclidean Space

The  $n$ -dimensional (real) *Euclidean space*  $\mathbb{R}^n$  is the collection of  $n$ -tuples of real numbers

$$\vec{x} = (x_1, x_2, \dots, x_n), \quad x_i \in \mathbb{R}.$$

Geometrically,  $\mathbb{R}^1$  is the usual real line and  $\mathbb{R}^2$  is a plane with origin. An element of  $\mathbb{R}^n$  can be considered as a *point*, or as an arrow starting from the origin and ending at the point, called a *vector*.

The *addition* and *scalar multiplication* of Euclidean vectors are

$$\vec{x} + \vec{y} = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n), \quad c\vec{x} = (cx_1, cx_2, \dots, cx_n).$$

They satisfy the usual properties such as commutativity and associativity. In general, a *vector space* is a set with two operations satisfying these usual properties.

The *dot product* of two Euclidean vectors is

$$\vec{x} \cdot \vec{y} = x_1y_1 + x_2y_2 + \dots + x_ny_n.$$

It satisfies the usual properties such as bilinearity and positivity. In general, an *inner product* on a vector space is a number valued operation satisfying these usual properties.

The dot product, or the inner product in general, induces the *Euclidean length*

$$\|\vec{x}\|_2 = \sqrt{\vec{x} \cdot \vec{x}} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}.$$

It also induces the *angle*  $\theta$  between two nonzero vectors by the formula

$$\cos \theta = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|_2 \|\vec{y}\|_2}.$$

The definition of the angle is justified by the *Schwarz's inequality*

$$|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\|_2 \|\vec{y}\|_2.$$

By the angle formula, two vectors are *orthogonal* and denoted  $\vec{x} \perp \vec{y}$ , if  $\vec{x} \cdot \vec{y} = 0$ . Moreover, the *orthogonal projection* of a vector  $\vec{x}$  on another vector  $\vec{y}$  is

$$\text{proj}_{\vec{y}} \vec{x} = \|\vec{x}\|_2 \cos \theta \frac{\vec{y}}{\|\vec{y}\|_2} = \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}} \vec{y}.$$

Finally, the area of the parallelogram formed by two vectors is

$$A(\vec{x}, \vec{y}) = \|\vec{x}\|_2 \|\vec{y}\|_2 |\sin \theta| = \sqrt{(\vec{x} \cdot \vec{x})(\vec{y} \cdot \vec{y}) - (\vec{x} \cdot \vec{y})^2}.$$

## Norm

The distance between numbers is defined by the absolute value. In a Euclidean space, however, the choice of absolute value is not unique.

**Definition 6.1.1.** A *norm* on a vector space is a function  $\|\vec{x}\|$  satisfying

1. *Positivity:*  $\|\vec{x}\| \geq 0$ , and  $\|\vec{x}\| = 0$  if and only if  $\vec{x} = \vec{0} = (0, 0, \dots, 0)$ .
2. *Scalar Property:*  $\|c\vec{x}\| = |c| \|\vec{x}\|$ .
3. *Triangle Inequality:*  $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ .

The norm induces the distance  $\|\vec{x} - \vec{y}\|$  between two vectors. Besides the Euclidean norm (length)  $\|\vec{x}\|_2$ , the other popular norms are

$$\begin{aligned} \|\vec{x}\|_1 &= |x_1| + |x_2| + \dots + |x_n|, \\ \|\vec{x}\|_\infty &= \max\{|x_1|, |x_2|, \dots, |x_n|\}. \end{aligned}$$

Exercise 6.2 defines the  $L^p$ -norm for any  $p \geq 1$ .

Any norm on  $\mathbb{R}$  is given by

$$\|x\| = \|x1\| = |x| \|1\| = c|x|, \quad c = \|1\|.$$

Here  $x1$  means the product of the scalar  $x$  to the vector  $1$ , and  $|x|$  is the usual absolute value. The equality shows that the norm on  $\mathbb{R}$  is unique up to multiplying a constant. This is why we did not need to discuss the norm for single variable sequences and functions.

Given any norm, we have the (*open*) *ball* and *closed ball* of radius  $\epsilon$  and centered at  $\vec{x}$

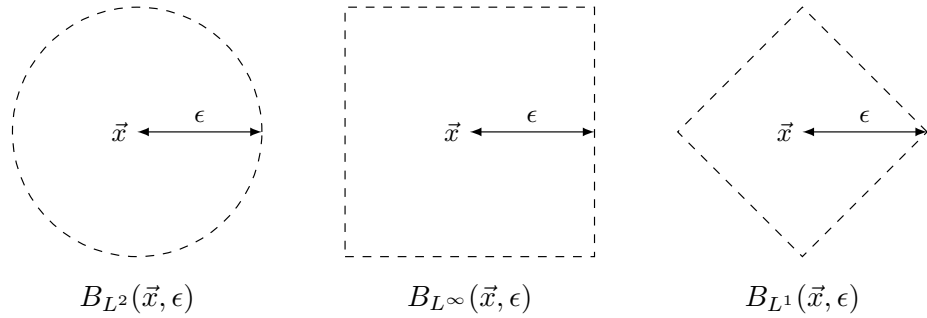
$$\begin{aligned} B(\vec{x}, \epsilon) &= \{\vec{y}: \|\vec{y} - \vec{x}\| < \epsilon\}, \\ \bar{B}(\vec{x}, \epsilon) &= \{\vec{y}: \|\vec{y} - \vec{x}\| \leq \epsilon\}. \end{aligned}$$

Moreover, for the Euclidean norm, we have the (closed) Euclidean ball and the sphere of radius  $R$

$$B_R^n = \{\vec{x}: \|\vec{x}\|_2 \leq R\} = \{(x_1, x_2, \dots, x_n): x_1^2 + x_2^2 + \dots + x_n^2 \leq R^2\},$$

$$S_R^{n-1} = \{\vec{x}: \|\vec{x}\|_2 = R\} = \{(x_1, x_2, \dots, x_n): x_1^2 + x_2^2 + \dots + x_n^2 = R^2\}.$$

When the radius  $R = 1$ , we get the *unit ball*  $B^n = B_1^n$  and the *unit sphere*  $S^{n-1} = S_1^{n-1}$ .



**Figure 6.1.1.** Balls with respect to different norms.

**Exercise 6.1.** Directly verify Schwarz's inequality for the dot product

$$|x_1y_1 + x_2y_2 + \dots + x_ny_n| \leq \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \sqrt{y_1^2 + y_2^2 + \dots + y_n^2},$$

and find the condition for the equality to hold. Moreover, use Schwarz's inequality to prove the triangle inequality for the Euclidean norm  $\|\vec{x}\|_2$ .

Note that Hölder's inequality in Exercise 3.75 generalizes Schwarz's inequality.

**Exercise 6.2.** Use Minkowski's inequality in Exercise 3.76 to prove that for any  $p \geq 1$ , the  $L^p$ -norm

$$\|\vec{x}\|_p = \sqrt[p]{|x_1|^p + |x_2|^p + \dots + |x_n|^p}$$

satisfies the three conditions for the norm. Moreover, prove that the  $L^p$ -norm satisfies

$$\|\vec{x}\|_\infty \leq \|\vec{x}\|_p \leq \sqrt[p]{n} \|\vec{x}\|_\infty.$$

**Exercise 6.3.** Let  $a_1, a_2, \dots, a_n$  be positive numbers and let  $p \geq 1$ . Prove that

$$\|\vec{x}\| = \sqrt[p]{a_1|x_1|^p + a_2|x_2|^p + \dots + a_n|x_n|^p}$$

is a norm.

**Exercise 6.4.** Suppose  $\|\cdot\|$  and  $|||\cdot|||$  are norms on  $\mathbb{R}^m$  and  $\mathbb{R}^n$ . Prove that  $\|(\vec{x}, \vec{y})\| = \max\{\|\vec{x}\|, |||\vec{y}|||\}$  is a norm on  $\mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^{m+n}$ . Prove that if  $m = n$ , then  $\|\vec{x}\| + |||\vec{x}|||$  is a norm on  $\mathbb{R}^n$ .

**Exercise 6.5.** Prove that for any norm and any vector  $\vec{x}$ , there is a number  $r \geq 0$  and a vector  $\vec{u}$ , such that  $\vec{x} = r\vec{u}$  and  $\|\vec{u}\| = 1$ . The expression  $\vec{x} = r\vec{u}$  is the *polar decomposition* that characterizes a vector by the length  $r$  and the direction  $\vec{u}$ .

**Exercise 6.6.** Prove that any norm satisfies  $||\vec{x}|| - ||\vec{y}|| \leq ||\vec{x} - \vec{y}||$ .

**Exercise 6.7.** Prove that if  $\vec{y} \in B(\vec{x}, \epsilon)$ , then  $B(\vec{y}, \delta) \subset B(\vec{x}, \epsilon)$  for some radius  $\delta > 0$ . In fact, we can take  $\delta = \epsilon - ||\vec{y} - \vec{x}||$ .

Norms can also be introduced in infinite dimensional vector spaces. For example, consider the vector space of all bounded sequences

$$l^\infty = \{\vec{x} = (x_1, x_2, \dots) : |x_i| \leq B \text{ for a constant } B \text{ and all } i\}.$$

It can be easily verified that

$$||\vec{x}||_\infty = \sup |x_i|$$

satisfies the three properties and is therefore a norm on  $V$ . The analysis on infinite dimensional vector spaces is the field of functional analysis.

**Exercise 6.8.** Let

$$l^2 = \{\vec{x} = (x_1, x_2, \dots) : \sum x_i^2 \text{ converges}\}.$$

Prove that  $l^2$  is a vector space, and

$$||\vec{x}||_2 = \sqrt{\sum x_i^2}$$

is a norm.

**Exercise 6.9.** For  $p \geq 1$ , introduce the infinite dimensional vector space  $l^p$  and the  $L^p$ -norm  $||\cdot||_p$ . Can you introduce a modified space and norm by using the idea of Exercise 6.3?

## Limit

The limit of a vector sequence can be defined with respect to a norm.

**Definition 6.1.2.** A sequence  $\vec{x}_k$  of vectors converges to a vector  $\vec{a}$  with respect to a norm  $||\vec{x}||$ , and denoted  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{a}$ , if for any  $\epsilon > 0$ , there is  $N$ , such that

$$k > N \implies ||\vec{x}_k - \vec{a}|| < \epsilon.$$

For the moment,  $m$  is used in the definition because  $n$  has been reserved for  $\mathbb{R}^n$ . The property  $||\vec{x}_k - \vec{a}|| < \epsilon$  is the same as  $\vec{x}_k \in B(\vec{a}, \epsilon)$ . The definition appears to depend on the choice of norm. For example,  $\lim_{k \rightarrow \infty} (x_k, y_k) = (a, b)$  with respect to the Euclidean norm means

$$\lim_{k \rightarrow \infty} \sqrt{|x_k - a|^2 + |y_k - b|^2} = 0,$$

and the same limit with respect to the  $L^\infty$ -norm means each coordinate converges

$$\lim_{k \rightarrow \infty} x_k = a, \quad \lim_{k \rightarrow \infty} y_k = b.$$

On the other hand, two norms  $\|\vec{x}\|$  and  $\|\vec{x}\|$  are *equivalent* if

$$c_1 \|\vec{x}\| \leq \|\vec{x}\| \leq c_2 \|\vec{x}\| \quad \text{for some constant } c_1, c_2 > 0.$$

It is easy to see that the convergence with respect to  $\|\vec{x}\|$  is equivalent to the convergence with respect to  $\|\vec{x}\|$ . For example, since all  $L^p$ -norms are equivalent by Exercise 6.2, the  $L^p$ -convergence in  $\mathbb{R}^n$  means exactly that each coordinate converges.

In Theorem 6.3.8, we will prove that all norms on  $\mathbb{R}^n$  (and on finite dimensional vector spaces in general) are equivalent. Therefore the convergence is in fact independent of the choice of the norm. If certain limit property (say, coordinate wise convergence) is proved only for the  $L^p$ -norm for the moment, then after Theorem 6.3.8 is established, it will become valid for all norms on finite dimensional vector spaces.

**Example 6.1.1.** We wish to prove that, if  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{a}$  and  $\lim_{k \rightarrow \infty} \vec{y}_k = \vec{b}$  with respect to a norm  $\|\cdot\|$ , then  $\lim_{k \rightarrow \infty} (\vec{x}_k + \vec{y}_k) = \vec{a} + \vec{b}$  with respect to the same norm.

We may simply copy the proof of Proposition 1.2.3, changing the absolute value to the norm. For any  $\epsilon > 0$ , there are  $N_1$  and  $N_2$ , such that

$$\begin{aligned} k > N_1 &\implies \|\vec{x}_k - \vec{a}\| < \frac{\epsilon}{2}, \\ k > N_2 &\implies \|\vec{y}_k - \vec{b}\| < \frac{\epsilon}{2}. \end{aligned}$$

Then for  $k > \max\{N_1, N_2\}$ , we have

$$\|(\vec{x}_k + \vec{y}_k) - (\vec{a} + \vec{b})\| \leq \|\vec{x}_k - \vec{a}\| + \|\vec{y}_k - \vec{b}\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

The first inequality is the triangle inequality.

Alternatively, we may prove the property by considering each coordinates. Let  $\vec{x}_k = (x_{1k}, \dots, x_{nk})$ ,  $\vec{y}_k = (y_{1k}, \dots, y_{nk})$ ,  $\vec{a} = (a_1, \dots, a_n)$ ,  $\vec{b} = (b_1, \dots, b_n)$ . Then with respect to the  $L^\infty$ -norm, the assumptions  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{a}$  and  $\lim_{k \rightarrow \infty} \vec{y}_k = \vec{b}$  mean that

$$\lim_{k \rightarrow \infty} x_{ik} = a_i, \quad \lim_{k \rightarrow \infty} y_{ik} = b_i.$$

By Proposition 1.2.3 (the arithmetic rule for single variable), this implies

$$\lim_{k \rightarrow \infty} (x_{ik} + y_{ik}) = a_i + b_i.$$

Then with respect to the  $L^\infty$ -norm, this means  $\lim_{k \rightarrow \infty} (\vec{x}_k + \vec{y}_k) = \vec{a} + \vec{b}$ .

The alternative proof only works for the  $L^\infty$ -norm (and therefore also for the  $L^p$ -norm). After Theorem 6.3.8, we know the the conclusion applies to any norm on  $\mathbb{R}^n$ .

**Exercise 6.10.** Prove that if the first norm is equivalent to the second norm, and the second norm is equivalent to the third norm, then the first norm is equivalent to the third norm.

**Exercise 6.11.** Prove that if  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{a}$  with respect to a norm  $\|\cdot\|$ , then  $\lim_{k \rightarrow \infty} \|\vec{x}_k\| = \|\vec{a}\|$ . Prove the converse is true if  $\vec{a} = \vec{0}$ . In particular, convergent sequences of vectors are bounded (extension of Proposition 1.2.1).

**Exercise 6.12.** Prove that if a sequence converges to  $\vec{a}$ , then any subsequence also converges to  $\vec{a}$ . This extends Proposition 1.2.2.

**Exercise 6.13.** Prove that if  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{a}$  and  $\lim_{k \rightarrow \infty} \vec{y}_k = \vec{b}$  with respect to a norm, then  $\lim_{k \rightarrow \infty} (\vec{x}_k + \vec{y}_k) = \vec{a} + \vec{b}$  with respect to the same norm.

**Exercise 6.14.** Prove that if  $\lim_{k \rightarrow \infty} c_k = c$  and  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{a}$  with respect to a norm, then  $\lim_{k \rightarrow \infty} c_k \vec{x}_k = c\vec{a}$  with respect to the same norm.

**Exercise 6.15.** Prove that if  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{a}$  and  $\lim_{k \rightarrow \infty} \vec{y}_k = \vec{b}$  with respect to the Euclidean norm  $\|\cdot\|_2$ , then  $\lim_{k \rightarrow \infty} \vec{x}_k \cdot \vec{y}_k = \vec{a} \cdot \vec{b}$ . Of course the Euclidean norm can be replaced by any norm after Theorem 6.3.8 is established.

The two proofs in Example 6.1.1 have different meanings. The first copies the proof for the single variable case (by replacing absolute value  $|\cdot|$  with norm  $\|\cdot\|$ ). Such proof uses only the three axioms for the norms, and is therefore valid for *all* norms, including norms on infinite dimensional vector spaces. The second proof uses convergence in individual coordinates, which only means the convergence with respect to the  $L^p$ -norm on  $\mathbb{R}^n$  (and any norm on finite dimensional vector space after Theorem 6.3.8 is established). Therefore such proof is not valid for general norms.

**Example 6.1.2.** The multivariable Bolzano-Weierstrass Theorem (Theorem 1.5.1) says that, if a sequence  $\vec{x}_k$  in  $\mathbb{R}^n$  is bounded with respect a norm  $\|\cdot\|$ , then the sequence has a converging subsequence (with respect to the same norm).

Consider the  $L^\infty$ -norm on  $\mathbb{R}^2$ . A sequence  $\vec{x}_k = (x_k, y_k)$  is  $L^\infty$ -bounded if and only if both coordinate sequences  $x_k$  and  $y_k$  are bounded. By Exercise 1.40, there are converging subsequences  $x_{k_p}$  and  $y_{k_p}$  *with the same indices*  $n_k$ . Then  $\vec{x}_{k_p} = (x_{k_p}, y_{k_p})$  is an  $L^\infty$ -convergent subsequence of  $\vec{x}_k$ .

Similar idea shows that the Bolzano-Weierstrass Theorem holds for the  $L^\infty$ -norm on any finite dimensional space  $\mathbb{R}^n$ . After Theorem 6.3.8 is established, we will know that the Bolzano-Weierstrass Theorem holds for any norm on any finite dimensional space.

The proof above uses individual coordinates, like the second proof in Example 6.1.1. In fact, we cannot prove Bolzano-Weierstrass Theorem like the first proof in Example 6.1.1, because the theorem fails for the infinite dimensional space  $l^\infty$  with the norm  $\|\cdot\|_\infty$ . Consider the sequence  $(1_{(k)})$  means that the  $k$ -th coordinate is 1)

$$\vec{x}_k = (0, 0, \dots, 0, 0, 1_{(k)}, 1, 1, \dots).$$

The sequence satisfies  $\|\vec{x}_k\|_\infty = 1$  and is therefore bounded. On the other hand, for  $k < l$  we always have

$$\|\vec{x}_k - \vec{x}_l\|_\infty = \|(0, \dots, 0, 1_{(k+1)}, \dots, 1_{(l)}, 0, 0, \dots)\|_\infty = 1.$$

This implies that any subsequence is not Cauchy and is therefore not convergent (see Exercise 6.17).

**Exercise 6.16.** If each coordinate of a sequence in  $l^\infty$  converges, can you conclude that the sequence converges with respect to the norm  $\|\cdot\|_\infty$ ? What about  $l^p$ ,  $p \geq 1$ ?

**Exercise 6.17.** Prove the easy direction of the Cauchy criterion: If  $\vec{x}_k$  converges with respect to a norm  $\|\cdot\|$ , then for any  $\epsilon > 0$ , there is  $N$ , such that

$$k, l > N \implies \|\vec{x}_k - \vec{x}_l\| < \epsilon.$$

**Exercise 6.18.** For the  $L^\infty$ -norm, prove that any Cauchy sequence converges. After Theorem 6.3.8 is established, we will know that the Cauchy criterion holds for any norm on any finite dimensional space.

**Exercise 6.19.** Prove that  $\vec{a}$  is the limit of a convergent subsequence of  $\vec{x}_k$  if and only if for any  $\epsilon > 0$ , there are infinitely many  $\vec{x}_k$  satisfying  $\|\vec{x}_k - \vec{a}\| < \epsilon$ . This extends Proposition 1.5.3 to any norm.

**Exercise 6.20.** Suppose  $\vec{x}_k$  does not converge to  $\vec{a}$ . Prove that there is a subsequence  $\vec{x}_{k_p}$ , such that every subsequence of  $\vec{x}_{k_p}$  does not converge to  $\vec{a}$ .

Because of the lack of order among vectors, Proposition 1.4.4 cannot be extended directly to vectors, but can still be applied to individual coordinates. For the same reason, the concept of upper and lower limits cannot be extended.

We can also extend the theory of series to vectors. We can use any norm in place of the absolute value in the discussion. Once we know the equivalence of all norms by Theorem 6.3.8, we will then find that the discussion is independent of the choice of norm. We can formulate the similar Cauchy criterion, which in particular implies that if  $\sum \vec{x}_k$  converges, then  $\lim \vec{x}_k = \vec{0}$ . We may also define the absolute convergence for series of vectors and extend Theorem 5.2.5. In fact, Theorems 5.2.5 and 5.2.6 on the rearrangements can be unified into the following remarkable theorem.

**Theorem 6.1.3 (Lévy-Steinitz).** *For any given series, the set of limits of all convergent rearrangements is either empty or a translation of a linear subspace.*

For a convergent series of vectors  $\sum \vec{x}_k$ , we may consider the “directions of absolute convergence”, which are vectors  $\vec{a}$ , such that  $\sum \vec{a} \cdot \vec{x}_k$  absolutely converges. Theorems 5.2.5 says that the limit of the series along such directions  $\vec{a}$  cannot be changed by rearrangements. Therefore the limit of the series can only be changed in directions orthogonal to such  $\vec{a}$ . Indeed, if  $A$  the collection of all such vectors  $\vec{a}$ , then the set of limits is the orthogonal complement  $A^\perp$  translated by the vector  $\sum \vec{x}_k$ . The proof of the theorem is out of the scope of this course<sup>32</sup>.

**Example 6.1.3.** A series  $\sum \vec{x}_k$  absolutely converges if  $\sum \|\vec{x}_k\|$  converges. We claim that the absolute convergence implies converges.

The convergence of means that, for any  $\epsilon > 0$ , there is  $N$ , such that

$$k \geq l > N \implies \|\vec{x}_l\| + \|\vec{x}_{l+1}\| + \cdots + \|\vec{x}_k\| < \epsilon.$$

<sup>32</sup>See “The Remarkable Theorem of Lévy and Steinitz” by Peter Rosenthal, American Math Monthly **94** (1987) 342-351.



By  $\|\vec{x}_l + \vec{x}_{l+1} + \cdots + \vec{x}_k\| \leq \|\vec{x}_l\| + \|\vec{x}_{l+1}\| + \cdots + \|\vec{x}_k\|$ , this gives

$$k \geq l > N \implies \|\vec{x}_l + \vec{x}_{l+1} + \cdots + \vec{x}_k\| < \epsilon.$$

By the Cauchy criterion (see Exercise 6.18), this implies that the series converges. The proof is the same as the proof of the comparison test (Proposition 5.2.1).

Note the use of Cauchy criterion here. For an infinite dimensional vector space, the Cauchy criterion may not hold. A norm is *complete* if the Cauchy criterion holds. After Theorem 6.3.8 is established, we will know that any finite dimensional normed space is complete.

**Exercise 6.21.** Prove the vector version of Dirichlet test (Proposition 5.2.3): Suppose  $a_k$  is a monotone sequence of numbers converging to 0. Suppose the partial sum of  $\sum \vec{x}_k$  is bounded. Then  $\sum a_k \vec{x}_k$  converges. Can you extend the Abel test in Proposition 5.2.4?

**Exercise 6.22.** Prove that the sum of absolutely convergent series of vectors is not changed by rearrangement. See Theorem 5.2.5.

**Exercise 6.23.** Suppose  $\sum a_k$  and  $\sum \vec{x}_k$  absolutely converge. Prove that  $\sum a_k \vec{x}_l$  absolutely converges, and  $\sum a_k \vec{x}_l = (\sum a_k)(\sum \vec{x}_k)$ .

## 6.2 Multivariable Map

In the study of single variable functions, we often assume the functions are defined on intervals. For multivariable functions and maps, it is no longer sufficient to only consider rectangles. In general, the domain of a multivariable map can be any subset

$$F(\vec{x}): A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m.$$

If  $m = 1$ , then the values of the map are real numbers, and  $F$  is a multivariable function. The coordinates of a multivariable map are multivariable functions

$$F(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \dots, f_m(\vec{x})).$$

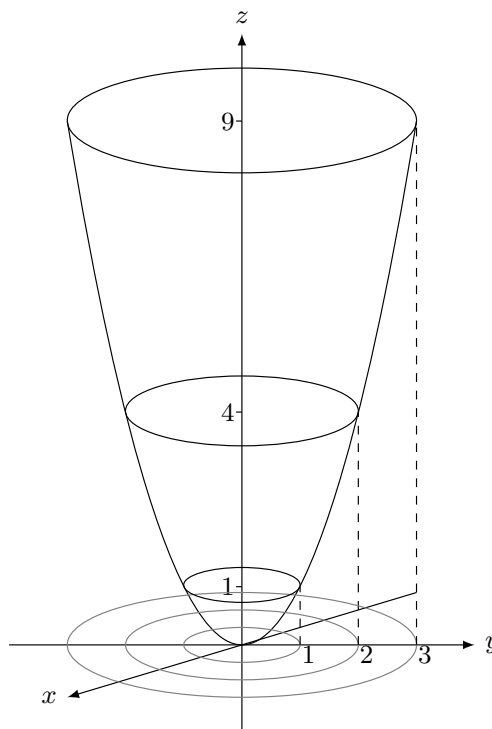
Two maps into the same Euclidean space may be added. A scalar number can be multiplied to a map into a Euclidean space. If  $F: A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $G: B \subset \mathbb{R}^m \rightarrow \mathbb{R}^k$  are maps such that  $F(\vec{x}) \in B$  for any  $\vec{x} \in A$ , then we have the composition  $(G \circ F)(\vec{x}) = G(F(\vec{x})): A \subset \mathbb{R}^n \rightarrow \mathbb{R}^k$ .

### Visualize Multivariable Map

Some special cases of maps can be visualized in various ways. For example, a multivariable function  $f$  on a subset  $A$  may be visualized either by the graph  $\{(\vec{x}, f(\vec{x})): \vec{x} \in A\}$  or the levels  $\{\vec{x} \in A: f(\vec{x}) = c\}$ .

A *parameterized curve (path)* in  $\mathbb{R}^n$  is a map

$$\phi(t) = (x_1(t), x_2(t), \dots, x_n(t)): [a, b] \rightarrow \mathbb{R}^n.$$



**Figure 6.2.1.** Graph and level of  $x^2 + y^2$ .

For example, the straight line passing through  $\vec{a}$  and  $\vec{b}$  is

$$\phi(t) = (1-t)\vec{a} + t\vec{b} = \vec{a} + t(\vec{b} - \vec{a}),$$

and the unit circle on the plane is

$$\phi(t) = (\cos t, \sin t): [0, 2\pi] \rightarrow \mathbb{R}^2.$$

Usually we require each coordinate function  $x_i(t)$  to be continuous. We will see that this is equivalent to the continuity of the map  $\phi$ . We say the curve connects  $\phi(a)$  to  $\phi(b)$ . A subset  $A \subset \mathbb{R}^n$  is *path connected* if any two points in  $A$  are connected by a curve in  $A$ .

Similar to curves, a map  $\mathbb{R}^2 \rightarrow \mathbb{R}^n$  may be considered as a *parameterized surface*. For example, the sphere in  $\mathbb{R}^3$  may be parameterized by

$$\sigma(\phi, \theta) = (\sin \phi \cos \theta, \sin \phi \sin \theta, \cos \phi): [0, \pi] \times [0, 2\pi] \rightarrow \mathbb{R}^3, \quad (6.2.1)$$

and the torus by ( $a > b > 0$ )

$$\sigma(\phi, \theta) = ((a+b \cos \phi) \cos \theta, (a+b \cos \phi) \sin \theta, b \sin \phi): [0, 2\pi] \times [0, 2\pi] \rightarrow \mathbb{R}^3. \quad (6.2.2)$$

A map  $\mathbb{R}^n \rightarrow \mathbb{R}^n$  may be considered as a *change of variable*, or a *transform*, or a *vector field*. For example,

$$(x, y) = (r \cos \theta, r \sin \theta): [0, +\infty) \times \mathbb{R} \rightarrow \mathbb{R}^2$$

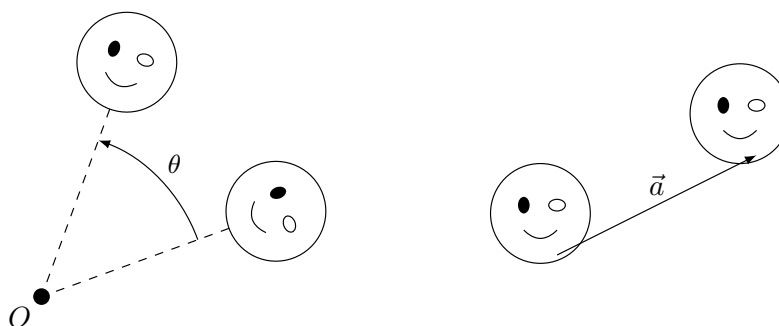
is the change of variables between the cartesian coordinate and the polar coordinate. Moreover,

$$R_\theta(x, y) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta): \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

transforms the plane by rotation of angle  $\theta$ , and

$$\vec{x} \mapsto \vec{a} + \vec{x}: \mathbb{R}^n \rightarrow \mathbb{R}^n$$

shifts the whole Euclidean space by  $\vec{a}$ .



**Figure 6.2.2.** *Rotation and shifting.*

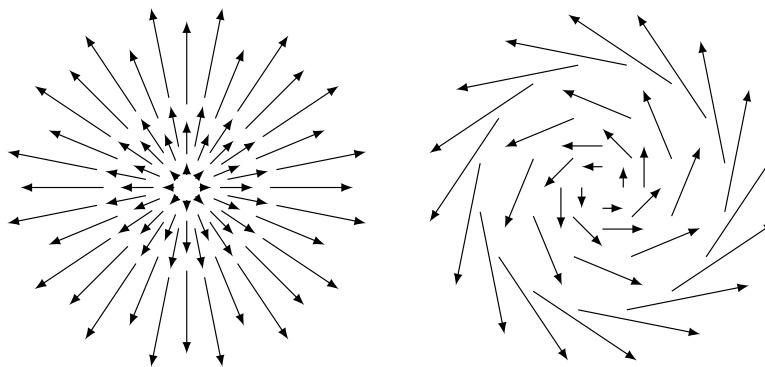
A vector field assigns an arrow to a point. For example, the map

$$F(\vec{x}) = \vec{x}: \mathbb{R}^n \rightarrow \mathbb{R}^n$$

is a vector field in the radial direction, while

$$F(x, y) = (y, -x): \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

is a counterclockwise rotating vector field, just like the water flow in the sink.



**Figure 6.2.3.** *Radial and rotational vector fields.*

**Exercise 6.24.** Describe the graph and level of function.

- |                      |  |  |
|----------------------|--|--|
| 1. $ax + by$ .       | 4. $ x ^p +  y ^p$ .                                     | 7. $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}$ . |
| 2. $ax + by + cz$ .  | 5. $xy$ .  |  |
| 3. $(x^2 + y^2)^2$ . | 6. $\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2}$ . | 8. $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2}$ . |

**Exercise 6.25.** Describe the maps in suitable ways. The meaning of the sixth map can be seen through the polar coordinate.

- |                                   |                                       |
|-----------------------------------|---------------------------------------|
| 1. $F(x) = (\cos x, \sin x, x)$ . | 5. $F(x, y) = (x^2, y^2)$ .           |
| 2. $F(x, y) = (x, y, x + y)$ .    | 6. $F(x, y) = (x^2 - y^2, 2xy)$ .     |
| 3. $F(x, y, z) = (y, z, x)$ .     | 7. $F(\vec{x}) = 2\vec{x}$ .          |
| 4. $F(x, y, z) = (y, -x, z)$ .    | 8. $F(\vec{x}) = \vec{a} - \vec{x}$ . |

## Limit

**Definition 6.2.1.** A multivariable map  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  has *limit*  $\vec{l}$  at  $\vec{a}$  with respect to given norms on  $\mathbb{R}^n$  and  $\mathbb{R}^m$ , and denoted  $\lim_{\vec{x} \rightarrow \vec{a}} f(\vec{x}) = \vec{l}$ , if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$0 < \|\vec{x} - \vec{a}\| < \delta \implies \|F(\vec{x}) - \vec{l}\| < \epsilon.$$

To emphasize that the map is defined on a subset  $A$ , we sometimes denote the limit by  $\lim_{\vec{x} \in A, \vec{x} \rightarrow \vec{a}} F(\vec{x}) = \vec{l}$ . If  $B \subset A$ , then by restricting the definition of the limit from  $\vec{x} \in A$  to  $\vec{x} \in B$ , we find

$$\lim_{\vec{x} \in A, \vec{x} \rightarrow \vec{a}} F(\vec{x}) = \vec{l} \implies \lim_{\vec{x} \in B, \vec{x} \rightarrow \vec{a}} F(\vec{x}) = \vec{l}.$$

In particular, if the restrictions of a map on different subsets give different limits, then the limit diverges. On the other hand, for a finite union,  $\lim_{\vec{x} \in A_1 \cup \dots \cup A_k, \vec{x} \rightarrow \vec{a}} F(\vec{x})$  converges if and only if all  $\lim_{\vec{x} \in A_i, \vec{x} \rightarrow \vec{a}} F(\vec{x})$  converge to the same limit value.

Let  $f_i$  and  $l_i$  be the coordinates of  $F$  and  $\vec{l}$ . For the  $L^\infty$ -norm on  $\mathbb{R}^m$ , it is easy to see that

$$\lim_{\vec{x} \rightarrow \vec{a}} F(\vec{x}) = \vec{l} \iff \lim_{\vec{x} \rightarrow \vec{a}} f_i(\vec{x}) = l_i \text{ for all } i.$$

Once we prove all norms are equivalent in Theorem 6.3.8, the property also holds for any norm on  $\mathbb{R}^m$ .

The limit of multivariable maps and functions has most of the usual properties enjoyed by single variable functions. The major difference between functions and maps is that values of functions can be compared, but values of maps cannot be compared. For example, the following extends the similar Propositions 2.1.7 and 2.3.3 for single variable functions ( $k$  is used because  $m$  and  $n$  are reserved for dimensions). The same proof applies here.

**Proposition 6.2.2.** For a map  $F$ ,  $\lim_{\vec{x} \rightarrow \vec{a}} F(\vec{x}) = \vec{l}$  if and only if  $\lim_{n \rightarrow \infty} \vec{x}_k = \vec{a}$  and  $\vec{x}_k \neq \vec{a}$  imply  $\lim_{k \rightarrow \infty} F(\vec{x}_k) = \vec{l}$ .

We may also define various variations at  $\infty$  by replacing  $\|\vec{x} - \vec{a}\| < \delta$  with  $\|\vec{x}\| > N$  or replacing  $|f(\vec{x}) - l| < \epsilon$  by  $|f(\vec{x})| > b$ .

**Example 6.2.1.** Consider  $f(x, y) = \frac{xy(x^2 - y^2)}{x^2 + y^2}$  defined for  $(x, y) \neq (0, 0)$ . By  $|f(x, y)| \leq |xy|$ , we have  $\|(x, y)\|_\infty < \delta = \sqrt{\epsilon}$  implying  $|f(x, y)| \leq \epsilon$ . Therefore  $\lim_{x, y \rightarrow 0} f(x, y) = 0$  in the  $L^\infty$ -norm. By the equivalence between the  $L^p$ -norms, the limit also holds for other  $L^p$ -norms.

**Example 6.2.2.** Consider  $f(x, y) = \frac{xy}{x^2 + y^2}$  defined for  $(x, y) \neq (0, 0)$ . We have  $f(x, cx) = \frac{c}{1 + c^2}$ , so that  $\lim_{y=cx, (x, y) \rightarrow (0, 0)} f(x, y) = \frac{c}{1 + c^2}$ . Since the restriction of the function to straight lines of different slopes converge to different limits, the function diverges at  $(0, 0)$ .

**Example 6.2.3.** For any angle  $\theta \in [0, 2\pi)$ , we choose a number  $\delta_\theta > 0$ , and define

$$f(t \cos \theta, t \sin \theta) = \begin{cases} 0, & \text{if } |t| < \delta_\theta, \\ 1, & \text{if } |t| \geq \delta_\theta. \end{cases}$$

We also allow  $\delta_\theta = +\infty$ , for which we have  $f(t \cos \theta, t \sin \theta) = 0$  for all  $t$ . Then we have a function on  $\mathbb{R}^2$ , such that the limit at  $(0, 0)$  along any straight line is 0. On the other hand, by choosing

$$\delta_\theta = \begin{cases} \theta, & \text{if } \theta = \frac{1}{n}, \\ +\infty, & \text{otherwise,} \end{cases}$$

the corresponding function diverges at  $(0, 0)$ .

**Example 6.2.4.** Let  $f(x, y) = \frac{x^p y^q}{(x^m + y^n)^k}$ , with  $p, q, m, n, k > 0$ . We wish to find out when  $\lim_{x, y \rightarrow 0^+} f(x, y) = 0$ .

The limit we want is equivalent to the convergence to 0 when restricted to the following subsets

$$A = \{(x, y) : x^m \leq y^n, x > 0, y > 0\}, \quad B = \{(x, y) : y^n \leq x^m, x > 0, y > 0\}.$$

For  $(x, y) \in A$ , we have  $y^n \leq x^m + y^n \leq 2y^n$ , so that

$$2^{-k} x^p y^{q-nk} = \frac{x^p y^q}{(2y^n)^k} \leq \frac{x^p y^q}{(x^m + y^n)^k} \leq \frac{x^p y^q}{(y^n)^k} = x^p y^{q-nk}.$$

Therefore  $\lim_{(x, y) \in A, x, y \rightarrow 0^+} f(x, y) = 0$  if and only if  $\lim_{(x, y) \in A, x, y \rightarrow 0^+} x^p y^{q-nk} = 0$ . If we further restrict the limit to  $A \cap B = \{(x, y) : x^m = y^n, x > 0, y > 0\}$ , then we get the necessary condition  $\lim_{y \rightarrow 0^+} (y^{\frac{n}{m}})^p y^{q-nk} = 0$ , which means  $\frac{p}{m} + \frac{q}{n} > k$ . Conversely, we always have  $0 \leq x^p y^{q-nk} \leq (y^{\frac{n}{m}})^p y^{q-nk} = y^{n(\frac{p}{m} + \frac{q}{n} - k)}$  on  $A$ . Therefore  $\lim_{(x, y) \in A, x, y \rightarrow 0^+} x^p y^{q-nk} = 0$  if and only if  $\frac{p}{m} + \frac{q}{n} > k$ .

By the similar argument, we find  $\lim_{(x, y) \in B, x, y \rightarrow 0^+} f(x, y) = 0$  if and only if  $\frac{p}{m} + \frac{q}{n} > k$ . Thus we conclude that  $\lim_{x, y \rightarrow 0^+} f(x, y) = 0$  if and only if  $\frac{p}{m} + \frac{q}{n} > k$ .

Exercise 6.26. Find the condition for  $\lim_{x,y \rightarrow 0^+} f(x,y) = 0$ . Assume all the parameters are positive.

1.  $(x^p + y^q)^r \log(x^m + y^n)$ .
2.  $x^p \log\left(\frac{1}{x^m} + \frac{1}{y^n}\right)$ .
3.  $\frac{x^p y^q}{(x^m + y^n)^k (x^{m'} + y^{n'})^{k'}}$ .
4.  $\frac{(x^p + y^q)^r}{(x^m + y^n)^k}$ .

Exercise 6.27. Compute the convergent limits. All parameters are positive.

1.  $\lim_{(x,y) \rightarrow (1,1)} \frac{1}{x-y}$ .
2.  $\lim_{x,y,z \rightarrow 0} \frac{xyz}{x^2 + y^2 + z^2}$ .
3.  $\lim_{(x,y) \rightarrow (0,0)} (x-y) \sin \frac{1}{x^2 + y^2}$ .
4.  $\lim_{x,y \rightarrow \infty} (x-y) \sin \frac{1}{x^2 + y^2}$ .
5.  $\lim_{x,y \rightarrow \infty} \frac{(x^2 + y^2)^p}{(x^4 + y^4)^q}$ .
6.  $\lim_{0 < x < y^2, x \rightarrow 0, y \rightarrow 0} \frac{x^p y}{x^2 + y^2}$ .
7.  $\lim_{ax \leq y \leq bx, x, y \rightarrow \infty} \frac{1}{xy}$ .
8.  $\lim_{x,y \rightarrow 0^+} (x+y)^{xy}$ .
9.  $\lim_{x \rightarrow \infty, y \rightarrow 0} \left(1 + \frac{1}{x}\right)^{\frac{x^2}{x+y}}$ .
10.  $\lim_{x,y \rightarrow +\infty} \left(1 + \frac{1}{x}\right)^{\frac{x^2}{x+y}}$ .
11.  $\lim_{x,y \rightarrow +\infty} \frac{x^2 + y^2}{e^{x+y}}$ .

Exercise 6.28. Show that  $\lim_{x,y \rightarrow 0} \frac{xy^2}{x^2 + y^4}$  diverges, although the limit converges to 0 along any straight line leading to  $(0,0)$ .

Exercise 6.29. Extend Example 6.2.3 to higher dimension. For each vector  $\vec{v}$  in the unit sphere  $S^2$ , choose a number  $\delta_{\vec{v}} > 0$  and define

$$f(t\vec{v}) = \begin{cases} 0, & \text{if } |t| < \delta_{\vec{v}}, \\ 1, & \text{if } |t| \geq \delta_{\vec{v}}. \end{cases}$$

1. Find a sequence of vectors  $\vec{v}_n$  in  $S^2$ , such that no three vectors lie in the same 2-dimensional plane.
2. Choose  $\delta_{\vec{v}_n} = \frac{1}{n}$  and  $\delta_{\vec{v}} = +\infty$  otherwise. Prove that the corresponding  $f(x,y)$  diverges at  $\vec{0}$ , yet the restriction of the function to any plane converges to 0.

Exercise 6.30. Suppose  $f(x)$  and  $g(x)$  are defined on the right side of 0. Find the necessary and sufficient condition for  $\lim_{x,y \rightarrow 0^+} f(x)g(y)$  to converge.

Exercise 6.31. Suppose  $f(x)$  and  $g(y)$  are defined near 0. Find the necessary and sufficient condition for  $\lim_{x,y \rightarrow 0} f(x)g(y)$  to converge.

Exercise 6.32. Suppose  $\mathbb{R}^2$  has the  $L^\infty$ -norm. Prove that  $(f,g): \mathbb{R}^n \rightarrow \mathbb{R}^2$  is continuous if and only if both functions  $f$  and  $g$  are continuous.

Exercise 6.33. Prove Proposition 6.2.2.

### Repeated Limit

For multivariable maps, we may first take the limit in some variables and then take the limit in other variables. In this way, we get *repeated limits* such as  $\lim_{\vec{x} \rightarrow \vec{a}} \lim_{\vec{y} \rightarrow \vec{b}} F(\vec{x}, \vec{y})$ . The following result tells us the relation between the repeated limits and the usual limit  $\lim_{(\vec{x}, \vec{y}) \rightarrow (\vec{a}, \vec{b})} F(\vec{x}, \vec{y})$ . The proof is given for the norm  $\|(\vec{x}, \vec{y})\| = \max\{\|\vec{x}\|, \|\vec{y}\|\}$ .

**Proposition 6.2.3.** *Suppose  $F(\vec{x}, \vec{y})$  is defined near  $(\vec{a}, \vec{b})$ . Suppose*

$$\lim_{(\vec{x}, \vec{y}) \rightarrow (\vec{a}, \vec{b})} F(\vec{x}, \vec{y}) = \vec{l},$$

and

$$\lim_{\vec{y} \rightarrow \vec{b}} F(\vec{x}, \vec{y}) = G(\vec{x}) \text{ for each } \vec{x} \text{ near } \vec{a}.$$

Then

$$\lim_{\vec{x} \rightarrow \vec{a}} \lim_{\vec{y} \rightarrow \vec{b}} F(\vec{x}, \vec{y}) = \lim_{(\vec{x}, \vec{y}) \rightarrow (\vec{a}, \vec{b})} F(\vec{x}, \vec{y}).$$

*Proof.* For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$0 < \max\{\|\vec{x} - \vec{a}\|, \|\vec{y} - \vec{b}\|\} < \delta \implies \|F(\vec{x}, \vec{y}) - \vec{l}\| < \epsilon.$$

Then for each fixed  $\vec{x}$  near  $\vec{a}$ , taking  $\vec{y} \rightarrow \vec{b}$  in the inequality above gives us

$$0 < \|\vec{x} - \vec{a}\| < \delta \implies \|G(\vec{x}) - \vec{l}\| = \lim_{\vec{y} \rightarrow \vec{b}} \|F(\vec{x}, \vec{y}) - \vec{l}\| \leq \epsilon.$$

This implies that  $\lim_{\vec{x} \rightarrow \vec{a}} G(\vec{x}) = \vec{l}$ , and proves the proposition.  $\square$

**Example 6.2.5.** For  $f(x, y) = \frac{xy}{x^2 + y^2}$  in Example 6.2.2, we have

$$\lim_{x \rightarrow 0} f(x, y) = 0 \text{ for each } y, \quad \lim_{y \rightarrow 0} f(x, y) = 0 \text{ for each } x.$$

Therefore

$$\lim_{y \rightarrow 0} \lim_{x \rightarrow 0} f(x, y) = \lim_{x \rightarrow 0} \lim_{y \rightarrow 0} f(x, y) = 0.$$

However, the usual limit  $\lim_{(x, y) \rightarrow (0, 0)} f(x, y)$  diverges.

**Exercise 6.34.** Study  $\lim_{x \rightarrow 0, y \rightarrow 0}$ ,  $\lim_{x \rightarrow 0} \lim_{y \rightarrow 0}$  and  $\lim_{y \rightarrow 0} \lim_{x \rightarrow 0}$ . Assume  $p, q, r > 0$ .

$$1. \frac{x - y + x^2 + y^2}{x + y}.$$

$$3. \frac{|x|^p |y|^q}{(x^2 + y^2)^r}.$$

$$5. (x + y) \sin \frac{1}{x} \sin \frac{1}{y}.$$

$$2. x \sin \frac{1}{x} + y \cos \frac{1}{y}.$$

$$4. \frac{x^2 y^2}{|x|^3 + |y|^3}.$$

$$6. \frac{e^x - e^y}{\sin xy}.$$

**Exercise 6.35.** Define a concept of the uniform convergence and find the condition for properties such as  $\lim_{\vec{x} \rightarrow \vec{a}} \lim_{\vec{y} \rightarrow \vec{b}} F(\vec{x}, \vec{y}) = \lim_{\vec{y} \rightarrow \vec{b}} \lim_{\vec{x} \rightarrow \vec{a}} F(\vec{x}, \vec{y})$ , similar to Theorem 5.4.1.

**Exercise 6.36.** For each of three limits  $\lim_{(\vec{x}, \vec{y}) \rightarrow (\vec{a}, \vec{b})}$ ,  $\lim_{\vec{x} \rightarrow \vec{a}} \lim_{\vec{y} \rightarrow \vec{b}}$  and  $\lim_{\vec{y} \rightarrow \vec{b}} \lim_{\vec{x} \rightarrow \vec{a}}$ , construct an example so that the limit converges but the other two limits diverge. Moreover, for each limit, construct an example so that the limit diverges but the other two converge and are equal.

## Continuity

A map  $F$  defined on a subset  $A$  is *continuous* at  $\vec{a} \in A$  if  $\lim_{\vec{x} \rightarrow \vec{a}} F(\vec{x}) = F(\vec{a})$ . This means that, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\vec{x} \in A, \|\vec{x} - \vec{a}\| < \delta \implies \|F(\vec{x}) - F(\vec{a})\| < \epsilon.$$

The map is *uniformly continuous* on  $A$  if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\vec{x}, \vec{y} \in A, \|\vec{x} - \vec{y}\| < \delta \implies \|F(\vec{x}) - F(\vec{y})\| < \epsilon.$$

By Proposition 6.2.2, we know that  $F$  is continuous at  $\vec{a}$  if and only if the limit and the function commute

$$\lim_{n \rightarrow \infty} \vec{x}_k = \vec{a} \implies \lim_{n \rightarrow \infty} F(\vec{x}_k) = F(\vec{a}).$$

The continuity is also equivalent to the continuity of its coordinate functions (initially for  $L^\infty$ -norm, and eventually for all norms after Theorem 6.3.8). Many properties of continuous functions can be extended to maps. For example, the sum and the composition of continuous maps are still continuous. The scalar product of a continuous function and a continuous map is a continuous map. The dot product of two continuous maps is a continuous function.

**Example 6.2.6.** By direct argument, it is easy to see that the product function  $\mu(x, y) = xy: \mathbb{R}^2 \rightarrow \mathbb{R}$  is continuous. Suppose  $f(\vec{x})$  and  $g(\vec{x})$  are continuous. Then  $F(\vec{x}) = (f(\vec{x}), g(\vec{x})): \mathbb{R}^n \rightarrow \mathbb{R}^2$  is also continuous because each coordinate function is continuous. Therefore composition  $(\mu \circ F)(\vec{x}) = f(\vec{x})g(\vec{x})$  (which is the product of two functions) is also continuous.

**Exercise 6.37.** Prove that any norm is a continuous function with respect to itself. Is the norm a uniformly continuous function?

**Exercise 6.38.** Prove that  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is continuous with respect to the  $L^\infty$ -norm on  $\mathbb{R}^m$  if and only if  $F \cdot \vec{a}$  is a continuous function for any  $\vec{a} \in \mathbb{R}^m$ .

**Exercise 6.39.** Prove that if  $f(x, y)$  is monotone and continuous in  $x$  and is continuous in  $y$ , then  $f(x, y)$  is continuous with respect to the  $L^\infty$ -norm.

**Exercise 6.40.** Find a function  $f(x, y)$  that is continuous in  $x$  and in  $y$ , but is not continuous at a point  $(x_0, y_0)$  as a two variable function. Moreover, show that the two repeated limits  $\lim_{x \rightarrow x_0} \lim_{y \rightarrow y_0} f(x, y)$  and  $\lim_{y \rightarrow y_0} \lim_{x \rightarrow x_0} f(x, y)$  of such a function exist and are equal, yet the whole limit  $\lim_{(x, y) \rightarrow (x_0, y_0)} f(x, y)$  diverges.



**Exercise 6.41.** Prove that the addition  $(\vec{x}, \vec{y}) \rightarrow \vec{x} + \vec{y}: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  and the scalar multiplication  $(c, \vec{x}) \rightarrow c\vec{x}: \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  are continuous.

**Exercise 6.42.** Prove that if  $f(\vec{x})$  and  $g(\vec{x})$  are continuous at  $\vec{x}_0$ , and  $h(u, v)$  is continuous at  $(f(\vec{x}_0), g(\vec{x}_0))$  with respect to the  $L^\infty$ -norm on  $\mathbb{R}^2$ , then  $h(f(\vec{x}), g(\vec{x}))$  is continuous at  $\vec{x}_0$ .

**Exercise 6.43.** Prove that the composition of two continuous maps is continuous. Then use Exercise 6.41 to explain that the sum of two continuous maps is continuous. Moreover, please give a theoretical explanation to Exercise 6.42.

**Exercise 6.44.** Study the continuity.

$$1. \begin{cases} \frac{\sin xy}{x}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases} \quad 2. \begin{cases} y & \text{if } x \text{ is rational,} \\ -y, & \text{if } x \text{ is irrational.} \end{cases}$$

**Exercise 6.45.** Suppose  $f(x)$  and  $g(y)$  are functions defined near 0. Find the necessary and sufficient condition for  $f(x)g(y)$  to be continuous at  $(0, 0)$ .

**Exercise 6.46.** Suppose  $\phi(x)$  is uniformly continuous. Is  $f(x, y) = \phi(x)$  uniformly continuous with respect to the  $L^\infty$ -norm?

**Exercise 6.47.** A map  $F(\vec{x})$  is *Lipschitz* if  $\|F(\vec{x}) - F(\vec{y})\| \leq L\|\vec{x} - \vec{y}\|$  for some constant  $L$ . Prove that Lipschitz maps are uniformly continuous.

**Exercise 6.48.** Prove that the sum of uniformly continuous functions is uniformly continuous. What about the product?

**Exercise 6.49.** Prove that a uniformly continuous map on a bounded subset of  $\mathbb{R}^n$  is bounded.

**Exercise 6.50.** Prove that  $\lim_{\vec{x} \rightarrow \vec{a}} F(\vec{x})$  converges if and only if it converges along any continuous path leading to  $\vec{a}$ .

**Exercise 6.51.** Suppose  $f(x)$  is a continuous function on  $(a, b)$ . Then

$$g(x, y) = \frac{f(x) - f(y)}{x - y}$$

is a continuous function on  $(a, b) \times (a, b) - \Delta$ , where  $\Delta = \{(x, x) : x \in \mathbb{R}\}$  is the diagonal in  $\mathbb{R}^2$ .

1. Prove that if  $\lim_{(x,y) \rightarrow (c,c), x \neq y} g(x, y)$  converges, then  $f$  is differentiable at  $a$ .
2. Suppose  $f$  is differentiable near  $a$ . Prove that  $\lim_{(x,y) \rightarrow (c,c)} g(x, y)$  converges if and only if  $f'$  is continuous at  $c$ . In particular, this leads to a counterexample to the converse of the first part.
3. Prove that  $g$  extends to a continuous function on  $(a, b) \times (a, b)$  if and only if  $f$  is continuously differentiable on  $(a, b)$ .

4. Extend the discussion to the limit of

$$\frac{f(x) - f(y) - f'(y)(x - y)}{(x - y)^2}$$

at the diagonal.

## Intermediate Value Theorem

For single variable functions, the Intermediate Value Theorem says that the continuous image of any interval is again an interval. So the theorem is critically related to the interval. For multivariable functions, the interval may be replaced by path connected subset. Recall that a subset  $A$  is path connected if for any  $\vec{a}, \vec{b} \in A$ , there is a continuous map (i.e., a curve)  $\phi: [0, 1] \rightarrow A$ , such that  $\phi(0) = \vec{a}$  and  $\phi(1) = \vec{b}$ .

**Proposition 6.2.4.** *Suppose  $f(\vec{x})$  is a continuous function on a path connected subset  $A$ . Then for any  $\vec{a}, \vec{b} \in A$  and  $\gamma$  between  $f(\vec{a})$  and  $f(\vec{b})$ , there is  $\vec{c} \in A$ , such that  $f(\vec{c}) = \gamma$ .*

*Proof.* Since  $A$  is path connected, there is a continuous curve  $\phi(t): [0, 1] \rightarrow A$  such that  $\phi(0) = \vec{a}$  and  $\phi(1) = \vec{b}$ . The composition  $g(t) = f(\phi(t))$  is then a continuous function for  $t \in [0, 1]$ . Since  $\gamma$  is between  $g(0) = f(\vec{a})$  and  $g(1) = f(\vec{b})$ , by Theorem 2.5.1, there is  $t_0 \in [0, 1]$ , such that  $g(t_0) = \gamma$ . The conclusion is the same as  $f(\vec{c}) = \gamma$  for  $\vec{c} = \phi(t_0)$ .  $\square$

## 6.3 Compact Subset

We wish to extend properties of continuous single variable functions. For example, we will establish the following extension of Theorems 2.4.1 and 2.4.2.

**Theorem 6.3.1.** *A continuous function on a compact subset is bounded, uniformly continuous, and reaches its maximum and minimum.*

Needless to say, the compact subset is the higher dimensional version of bounded and closed intervals. The concept should be defined by certain properties so that the proofs of Theorems 2.4.1 and 2.4.2 remain valid. The following is the proof of the boundedness property, paraphrased from the original proof.

*Proof.* Suppose  $f$  is continuous on a compact subset  $K$ . If  $f$  is not bounded, then there is a sequence  $\vec{x}_k \in K$ , such that  $\lim_{k \rightarrow \infty} f(\vec{x}_k) = \infty$ . Since  $K$  is compact, there is a subsequence  $\vec{x}_{k_p}$  converging to  $\vec{a} \in K$ . By the continuity of  $f(\vec{x})$  at  $\vec{a}$ , this implies that  $f(\vec{x}_{k_p})$  converges to  $f(\vec{a})$ . Since this contradicts the assumption  $\lim_{k \rightarrow \infty} f(\vec{x}_k) = \infty$ , we conclude that  $f$  is actually bounded on  $K$ .  $\square$

Note that the proof is quite formal. It does not refer specifically to the norm, or even vectors in the Euclidean space. The proof actually makes sense in any topological space. From the analysis viewpoint, topology is the formal theory based on the formal notion of limits.

## Compact Subset

By examining the proof above, we get the following definition of compact subsets.

**Definition 6.3.2.** A subset of a Euclidean space is *compact* if any sequence in the subset has a convergent subsequence, and the limit of the subsequence still lies in the subset.

As remarked after the proof, a suitable generalization (called *net*, *Moore-Smith sequence*, or *filter*) of the definition is actually the concept of compact subset in any topological space. Theorem 6.3.1 can be extended to multivariable maps.

**Proposition 6.3.3.** *A continuous map  $F$  on a compact subset  $K$  is bounded and uniformly continuous. If  $F$  is a function, then it also reaches its maximum and minimum on  $K$ . Moreover, the image subset  $F(K) = \{F(\vec{x}) : \vec{x} \in K\}$  is also compact.*

A subset  $A \subset \mathbb{R}^n$  is *bounded* (with respect to a norm) if there is a constant  $B$ , such that  $\|\vec{x}\| < B$  for any  $\vec{x} \in A$ .

*Proof.* The proof of the usual properties of  $F$  is the same as before. We only prove the compactness of  $F(K)$ . A sequence in  $F(K)$  is  $F(\vec{x}_k)$ , with  $\vec{x}_k \in K$ . Since  $K$  is compact, there is a subsequence  $\vec{x}_{k_p}$  converging to  $\vec{a} \in K$ . By the continuity of  $F$  at  $\vec{a} \in K$ , we have

$$\lim_{k \rightarrow \infty} F(\vec{x}_{k_p}) = F(\vec{a}) \in F(K).$$

Therefore the subsequence  $F(\vec{x}_{k_p})$  of  $F(\vec{x}_k)$  converges to a limit in  $F(K)$ .  $\square$

**Example 6.3.1.** Let  $A = [a, b] \subset \mathbb{R}$  be a bounded and closed interval. Then any sequence  $x_k \in A$  is bounded. By Bolzano-Weierstrass Theorem (Theorem 1.5.1), there is a converging subsequence  $x_{k_p}$ . By  $a \leq x_{k_p} \leq b$  and the order rule, we have  $a \leq \lim x_{k_p} \leq b$ , which means  $\lim x_{k_p} \in A$ .

**Example 6.3.2.** The sequence  $k^{-1} \in A = \{k^{-1} : k \in \mathbb{N}\}$  converges to 0. Therefore any subsequence  $k_p^{-1}$  also converges to 0. This implies that no subsequence converges to a limit in  $A$ , and therefore  $A$  is not compact.

If we also include the limit 0, then we expect  $\bar{A} = A \cup \{0\}$  to be compact. Indeed, any sequence  $x_k \in \bar{A}$  is one of two kinds. The first is that infinitely many  $x_k$  are equal. This means that there is a constant subsequence  $x_{k_p} \in \bar{A}$ , which converges to  $x_{k_1} \in \bar{A}$ . The second is that there are infinitely many different  $x_k$ . For our specific  $\bar{A}$ , this implies that there is a subsequence  $x_{k_p} = \xi_p^{-1} \in A$  with  $\xi_p \rightarrow \infty$ . Then  $\lim x_{k_p} = 0 \in \bar{A}$ . We conclude that  $\bar{A}$  is compact.

**Example 6.3.3.** Suppose  $\lim \vec{x}_k = \vec{a}$ . Then  $F(k^{-1}) = \vec{x}_k$  and  $F(0) = \vec{a}$  define a map  $F : \bar{A} \rightarrow \mathbb{R}^n$  from the subset in Example 6.3.2. It is easy to verify that  $F$  is continuous. By Theorem 6.3.3, therefore, the image subset  $F(\bar{A})$  is also compact. Note that  $F(\bar{A})$  is the subset of all  $\vec{x}_k$  together with  $\vec{a}$ .

**Exercise 6.52.** With the help of Exercise 1.40, extend the argument in Example 6.3.1 to prove that any bounded and closed rectangle  $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$  is compact with respect to the  $L^\infty$ -norm.

**Exercise 6.53.** Suppose  $\vec{x}_n \in A$  converges to  $\vec{a} \notin A$ . Prove that  $A$  is not compact.

**Exercise 6.54.** Directly prove that the subset of all  $\vec{x}_n$  together with  $\vec{a}$  in Example 6.3.3 is a compact subset.

**Exercise 6.55.** Suppose  $K$  is a compact subset. Prove that there are  $\vec{a}, \vec{b} \in K$ , such that  $\|\vec{a} - \vec{b}\| = \max_{\vec{x}, \vec{y} \in K} \|\vec{x} - \vec{y}\|$ .

**Exercise 6.56.** Prove that the union of two compact subsets is compact.

**Exercise 6.57.** Using the product norm in Exercise 6.4, prove that if  $K, L$  are compact, then  $K \times L$  is compact. Conversely, if  $K, L \neq \emptyset$  and  $K \times L$  is compact, then  $K, L$  are compact.

**Exercise 6.58.** Prove the uniform continuity of a continuous map on a compact subset.

**Exercise 6.59.** Prove that a continuous function on a compact subset reaches its maximum and minimum. What is the meaning of the fact for the compact subset  $\bar{A}$  in Example 6.3.2?

## Inverse Map

We cannot talk about multivariable monotone maps. Thus the only part of Theorem 2.5.3 that can be extended is the continuity.

**Proposition 6.3.4.** *Suppose  $F: K \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a one-to-one and continuous map on a compact subset  $K$ . Then the inverse map  $F^{-1}: F(K) \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$  is also continuous.*

*Proof.* We claim that  $F^{-1}$  is in fact uniformly continuous. Suppose it is not uniformly continuous. Then there is  $\epsilon > 0$  and  $\vec{\xi}_k = F(\vec{x}_k), \vec{\eta}_k = F(\vec{y}_k) \in F(K)$ , such that

$$\lim_{k \rightarrow \infty} \|\vec{\xi}_k - \vec{\eta}_k\| = 0, \quad \|\vec{x}_k - \vec{y}_k\| = \|F^{-1}(\vec{\xi}_k) - F^{-1}(\vec{\eta}_k)\| \geq \epsilon.$$

By the compactness of  $K$ , we can find  $k_p$ , such that both  $\lim_{p \rightarrow \infty} \vec{x}_{k_p} = \vec{a}$  and  $\lim_{p \rightarrow \infty} \vec{y}_{k_p} = \vec{b}$  converge, and both  $\vec{a}$  and  $\vec{b}$  lie in  $K$ . Then by the continuity of  $F$  at  $\vec{a}$  and  $\vec{b}$ , we have

$$\lim_{p \rightarrow \infty} \vec{\xi}_{k_p} = \lim_{p \rightarrow \infty} F(\vec{x}_{k_p}) = F(\vec{a}), \quad \lim_{p \rightarrow \infty} \vec{\eta}_{k_p} = \lim_{p \rightarrow \infty} F(\vec{y}_{k_p}) = F(\vec{b}).$$

We conclude that  $\|F(\vec{a}) - F(\vec{b})\| = \lim_{p \rightarrow \infty} \|\vec{\xi}_{k_p} - \vec{\eta}_{k_p}\| = 0$ , so that  $F(\vec{a}) = F(\vec{b})$ . Since  $F$  is one-to-one, we get  $\vec{a} = \vec{b}$ . This contradicts with  $\lim_{p \rightarrow \infty} \vec{x}_{k_p} = \vec{a}$ ,  $\lim_{p \rightarrow \infty} \vec{y}_{k_p} = \vec{b}$ , and  $\|\vec{x}_k - \vec{y}_k\| \geq \epsilon$ .  $\square$

### Closed Subset

The concept of compact subsets is introduced as a high dimensional substitute of bounded and closed intervals. So we expect the compactness to be closely related to the high dimensional version of bounded and closed subsets.

By applying Bolzano-Weierstrass Theorem to each coordinate (see Exercise 1.40), we find that any sequence in a bounded subset  $K$  has a convergent subsequence (with respect to the  $L^\infty$ -norm). For  $K$  to be compact, therefore, it remains to satisfy the following

$$\vec{x}_k \in K, \lim_{k \rightarrow \infty} \vec{x}_k = \vec{l} \implies \vec{l} \in K.$$

**Definition 6.3.5.** A subset of a Euclidean space is *closed* if the limit of any convergent sequence in the subset still lies in the subset.

**Proposition 6.3.6.** *A subset of the Euclidean space is compact if and only if it is bounded and closed.*

*Proof.* For the special case of the  $L^\infty$ -norm, we have argued that bounded and closed subsets are compact. The converse will be proved for any norm.

By Exercise 6.37, the norm is a continuous function. By Proposition 6.3.3, the norm function is bounded on the compact subset. This means that the compact subset is bounded.

To prove that a compact subset  $K$  is closed, we consider a sequence  $\vec{x}_k \in K$  converging to  $\vec{l}$ , and would like to argue that  $\vec{l} \in K$ . By compactness, we have a subsequence  $\vec{x}_{k_p}$  converging to  $\vec{l} \in K$ . Then  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{l}$  implies the subsequence also converges to  $\vec{l}$ . By the uniqueness of limit, we get  $\vec{l} = \vec{l} \in K$ .  $\square$

**Proposition 6.3.7.** *Closed subsets have the following properties.*

1.  $\emptyset$  and  $\mathbb{R}^n$  are closed.
2. Intersections of closed subsets are closed.
3. Finite unions of closed subsets are closed.

*Proof.* The first property is trivially true.

Suppose  $C_i$  are closed and  $C = \cap C_i$ . Suppose  $\vec{x}_k \in C$  and  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{l}$ . Then the sequence  $\vec{x}_k$  lies in each  $C_i$ . Because  $C_i$  is closed, the limit  $\vec{l}$  lies in each  $C_i$ . Therefore  $\vec{l} \in \cap C_i = C$ .

Suppose  $C$  and  $D$  are closed. Suppose  $\vec{x}_k \in C \cup D$  and  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{l}$ . Then there must be infinitely many  $\vec{x}_k$  in either  $C$  or  $D$ . In other words, there is a subsequence in either  $C$  or  $D$ . If  $C$  contains a subsequence  $\vec{x}_{k_p}$ , then because  $C$  is closed, we get  $\vec{l} = \lim_{k \rightarrow \infty} \vec{x}_k = \lim_{p \rightarrow \infty} \vec{x}_{k_p} \in C$ . If  $D$  contains a subsequence, then we similarly get  $\vec{l} \in D$ . Therefore  $\vec{l} \in C \cup D$ .  $\square$

**Example 6.3.4.** Any closed interval  $[a, b]$  is a closed subset of  $\mathbb{R}$ . The reason is the order rule. Suppose  $x_k \in [a, b]$  and  $\lim x_k = l$ . By  $a \leq x_k \leq b$  and the order rule, we have  $a \leq l \leq b$ , which means  $l \in [a, b]$ .

**Exercise 6.60.** Prove that if two norms are equivalent, then a subset is bounded, compact, or closed with respect to one norm if and only if it has the same property with respect to the other norm.

**Exercise 6.61.** Prove that any closed rectangle  $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$  is a closed subset of  $\mathbb{R}^n$  with respect to the  $L^\infty$ -norm.

**Exercise 6.62.** Prove that both the closed ball  $\bar{B}(\vec{x}, r) = \{\vec{x}: \|\vec{x}\| \leq r\}$  and the sphere  $S(\vec{x}, r) = \{\vec{x}: \|\vec{x}\| = r\}$  are bounded and closed. Moreover, for the  $L^\infty$ -norm, prove that both  $\bar{B}(\vec{x}, r)$  and  $S(\vec{x}, r)$  are compact.

**Exercise 6.63.** Prove that any closed subset is union of countably many compact subsets.

**Exercise 6.64.** Using the product norm in Exercise 6.4, prove that if  $A, B$  are closed, then  $A \times B$  is closed. Conversely, if  $A, B \neq \emptyset$  and  $A \times B$  is closed, then  $A, B$  are closed.

**Exercise 6.65.** Suppose  $K$  is compact and  $C$  is closed. Prove that if  $C \subset K$ , then  $C$  is also compact.

**Exercise 6.66.** Suppose  $\vec{x}_k$  is a sequence with no converging subsequence. Prove that the subset of all  $\vec{x}_k$  is closed.

**Exercise 6.67.** Theorem 6.3.3 tells us the image of a bounded closed subset under a continuous map is still bounded and closed. Show that the image of a closed subset under a continuous map is not necessarily closed.

**Exercise 6.68.** Suppose a map  $F: \mathbb{R}^m \rightarrow \mathbb{R}^n$  is continuous and  $C$  is closed. Prove that the preimage  $F^{-1}(C) = \{\vec{x}: F(\vec{x}) \in C\}$  is closed.

**Exercise 6.69.** A collection of subsets  $A_i$  is *locally finite* if for any  $\vec{x} \in \mathbb{R}^n$ , there is  $\epsilon > 0$ , such that  $A_i \cap B(\vec{x}, \epsilon) = \emptyset$  for all but finitely many  $A_i$ . Prove that the union of a locally finite collection of closed subsets is also closed.

**Exercise 6.70.** For any subset  $A$  of  $\mathbb{R}^n$ , the *closure* of  $A$  is (compare with the definition of  $\text{LIM}\{x_k\}$  in Section 1.5)

$$\bar{A} = \left\{ \lim_{k \rightarrow \infty} \vec{x}_k : \vec{x}_k \in A \text{ and } \lim_{k \rightarrow \infty} \vec{x}_k \text{ converges} \right\}.$$

Prove that the closure is closed (compare with Exercise 1.47).

**Exercise 6.71.** Prove that the closure  $\bar{A}$  is the smallest closed subset containing  $A$ , and  $A$  is closed if and only if  $\bar{A} = A$ .

Exercise 6.72. Let  $\vec{x}_k$  be a sequence. Let

$$\text{LIM}\{\vec{x}_k\} = \{\lim \vec{x}_{k_p} : \vec{x}_{k_p} \text{ converges}\}.$$

1. Extend Proposition 1.5.3 to  $\text{LIM}\{\vec{x}_k\}$ , and prove that  $\lim \vec{x}_k$  converges if and only if  $\text{LIM}\{\vec{x}_k\}$  consists of single point.
2. Let  $A = \{\vec{x}_k\}$  be the collection of all terms in the sequence. Prove that  $\bar{A} = A \cup \text{LIM}\{\vec{x}_k\}$ .

Exercise 6.73. Prove the converse of Exercise 6.68: If a map  $F: \mathbb{R}^m \rightarrow \mathbb{R}^n$  has the property that the preimage of any closed subset is closed, then the map is continuous.

Suppose  $\lim \vec{x}_k = \vec{a}$  and  $\lim F(\vec{x}_k) \neq F(\vec{a})$ . First show that we may assume no subsequence of  $F(\vec{x}_k)$  converges to  $F(\vec{a})$ . Then apply the property to the closure of the collection of all the terms  $F(\vec{x}_k)$ .

## Equivalence of Norms

The following result allows us to freely use any norm in multivariable analysis.

**Theorem 6.3.8.** *All norms on a finite dimensional vector space are equivalent.*

*Proof.* We prove the claim for norms on a Euclidean space. Since linear algebra tells us that any finite dimensional vector space is *isomorphic* to a Euclidean space, the result applies to any finite dimensional vector space.

We compare any norm  $\|\vec{x}\|$  with the  $L^\infty$ -norm  $\|\vec{x}\|_\infty$ . Let  $\vec{e}_1 = (1, 0, \dots, 0)$ ,  $\vec{e}_2 = (0, 1, \dots, 0)$ ,  $\dots$ ,  $\vec{e}_n = (0, 0, \dots, 1)$  be the *standard basis* of  $\mathbb{R}^n$ . By the definition of norm, we have

$$\begin{aligned} \|\vec{x}\| &= \|x_1\vec{e}_1 + x_2\vec{e}_2 + \dots + x_n\vec{e}_n\| \\ &\leq |x_1|\|\vec{e}_1\| + |x_2|\|\vec{e}_2\| + \dots + |x_n|\|\vec{e}_n\| \\ &\leq (\|\vec{e}_1\| + \|\vec{e}_2\| + \dots + \|\vec{e}_n\|) \max\{|x_1|, |x_2|, \dots, |x_n|\} \\ &= (\|\vec{e}_1\| + \|\vec{e}_2\| + \dots + \|\vec{e}_n\|) \|\vec{x}\|_\infty = c_1 \|\vec{x}\|_\infty. \end{aligned}$$

It remains to prove that  $\|\vec{x}\| \geq c_2 \|\vec{x}\|_\infty$  for some  $c_2 > 0$ . A special case is

$$\|\vec{x}\|_\infty = 1 \implies \|\vec{x}\| \geq c_2.$$

Conversely, given the special case, we write any vector  $\vec{x}$  as  $\vec{x} = r\vec{u}$  with  $r = \|\vec{x}\|_\infty$  and  $\|\vec{u}\|_\infty = 1$  (see Exercise 6.5). Then we get the general case

$$\|\vec{x}\| = \|r\vec{u}\| = r\|\vec{u}\| \geq rc_2 = c_2 \|\vec{x}\|_\infty.$$

Therefore it is sufficient to prove the special case.

We interpret the special case as the norm function  $\|\vec{x}\|$  having a positive minimum on the subset  $K = \{\vec{x} : \|\vec{x}\|_\infty = 1\}$ . We know that  $K$  is bounded and closed with respect to the  $L^\infty$ -norm, so by Proposition 6.3.6 (fully proved for the  $L^\infty$ -norm),  $K$  is compact with respect to the  $L^\infty$ -norm (see Exercise 6.62). On the

other hand, the inequality  $|\|\vec{x}\| - \|\vec{y}\|| \leq \|\vec{x} - \vec{y}\| \leq c_1 \|\vec{x} - \vec{y}\|_\infty$  implies that the norm function  $\|\vec{x}\|$  is continuous with respect to the  $L^\infty$ -norm. Then by Theorem 6.3.1, the norm  $\|\vec{x}\|$  reaches its minimum on  $K$  at some  $\vec{a} \in K$ . The minimum value  $\|\vec{a}\|$  (to be taken as  $c_2$ ) is positive because

$$\vec{a} \in K \implies \|\vec{a}\|_\infty = 1 \implies \vec{a} \neq \vec{0} \implies \|\vec{a}\| > 0. \quad \square$$

## 6.4 Open Subset

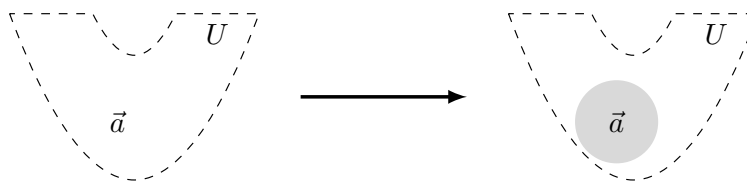
### Open Subset

In the analysis of single variable functions, such as differentiability, we often require functions to be defined at all the points near a given point. For multivariable functions, the requirement becomes the following concept.

**Definition 6.4.1.** A subset of a Euclidean space is *open* if for any point in the subset, all the points near the point are also in the subset.

Thus a subset  $U$  is open if

$$\vec{a} \in U \implies B(\vec{a}, \epsilon) \subset U \text{ for some } \epsilon > 0.$$



**Figure 6.4.1.** Definition of open subset.

**Exercise 6.74.** Prove that if two norms are equivalent, then a subset is open with respect to one norm if and only if it is open with respect to the other norm.

**Exercise 6.75.** Using the product norm in Exercise 6.4, prove that if  $A, B$  are open, then  $A \times B$  is open. Conversely, if  $A, B \neq \emptyset$  and  $A \times B$  is open, then  $A, B$  are open.

**Exercise 6.76.** The *characteristic function* of a subset  $A \subset \mathbb{R}^n$  is

$$\chi_A(\vec{x}) = \begin{cases} 1, & \text{if } \vec{x} \in A, \\ 0, & \text{if } \vec{x} \notin A. \end{cases}$$

Prove that  $\chi_A$  is continuous at  $\vec{a}$  if and only if some ball around  $\vec{a}$  is either contained in  $A$  or is disjoint from  $A$ .

**Exercise 6.77.** For a continuous function  $f$  on  $\mathbb{R}$ , prove that the subset  $\{(x, y) : y < f(x)\}$  is open. What if  $f$  is not continuous?



Closed subsets are defined by using sequence limit. The following is the interpretation of open subsets by sequence limit.

**Proposition 6.4.2.** *A subset  $U$  of Euclidean space is open if and only if for any sequence  $\vec{x}_k$  converging to a limit in  $U$ , we have  $\vec{x}_k \in U$  for sufficiently large  $k$ .*

*Proof.* Suppose  $U$  is open and  $\lim_{k \rightarrow \infty} \vec{x}_k = \vec{a} \in U$ . Then by the definition of  $U$  open, a ball  $B(\vec{a}, \epsilon)$  is contained in  $U$ . Moreover, by the definition of sequence limit, we have  $\|\vec{x}_k - \vec{a}\| < \epsilon$  for sufficiently large  $k$ . Then we get  $\vec{x}_k \in B(\vec{a}, \epsilon) \subset U$  for sufficiently large  $k$ .

Suppose  $U$  is not open. Then there is  $\vec{a}$ , such that  $B(\vec{a}, \epsilon)$  is never contained in  $U$ . By taking  $\epsilon = \frac{1}{k}$ , we get a sequence  $\vec{x}_k \notin U$  satisfying  $\|\vec{x}_k - \vec{a}\| < \frac{1}{k}$ . The sequence converges to  $\vec{a} \in U$  but completely lies outside  $U$ . The contradiction proves the converse.  $\square$

Now we can use the sequence limit to argue the following property.

**Proposition 6.4.3.** *A subset  $U$  of  $\mathbb{R}^n$  is open if and only if the complement  $\mathbb{R}^n - U$  is closed.*

*Proof.* Suppose  $U$  be open. To prove that the complement  $C = \mathbb{R}^n - U$  is closed, we assume that a sequence  $\vec{x}_k \in C$  converges to  $\vec{a}$  and wish to prove  $\vec{a} \in C$ . If  $\vec{a} \notin C$ , then  $\vec{a} \in U$ . By  $\vec{x}_k$  converging to  $\vec{a}$  and Proposition 6.4.2, this implies that  $\vec{x}_k \in U$  for sufficiently large  $k$ . Since this contradicts to the assumption  $\vec{x}_k \in C$ , we conclude that  $\vec{a} \in C$ .

Suppose  $C$  is closed. To prove that the complement  $C = \mathbb{R}^n - U$  is open, we assume that a sequence  $\vec{x}_k$  converges to  $\vec{a} \in U$  and (by Proposition 6.4.2) wish to prove  $\vec{x}_k \in U$  for sufficiently large  $k$ . If our wish is wrong, then there is a subsequence  $\vec{x}_{k_p} \notin U$ . Since  $\vec{x}_k$  converges to  $\vec{a}$ , the subsequence also converges to  $\vec{a}$ . Since  $\vec{x}_{k_p} \notin U$  means  $\vec{x}_{k_p} \in C$ , by  $C$  closed, the limit  $\vec{a}$  of the subsequence must also lie in  $C$ . This contradicts to the assumption  $\vec{a} \in U$ . Therefore we must have  $\vec{x}_k \in U$  for sufficiently large  $k$ .  $\square$

Combining Propositions 6.3.7 and 6.4.3 gives properties of open subsets.

**Proposition 6.4.4.** *Open subsets have the following properties.*

1.  $\emptyset$  and  $\mathbb{R}^n$  are open.
2. Unions of open subsets are open.
3. Finite intersections of open subsets are open.

The three properties can be used as the axiom of general topology theory.

**Exercise 6.78.** Prove Proposition 6.4.4 by directly using the definition of open subsets.

**Exercise 6.79.** Prove Proposition 6.4.4 by using the sequence limit characterisation of open subsets in Proposition 6.4.2.

The following describes the relation between open subsets and balls.

**Proposition 6.4.5.** *A subset is open if and only if it is a union of balls. In fact, we can choose the union to be countable.*

*Proof.* First we prove that any ball  $B(\vec{a}, \epsilon)$  is open. If  $\vec{x} \in B(\vec{a}, \epsilon)$ , then  $\|\vec{x} - \vec{a}\| < \epsilon$ . For  $\delta = \epsilon - \|\vec{x} - \vec{a}\| > 0$ , we have  $B(\vec{x}, \delta) \subset B(\vec{a}, \epsilon)$  by

$$\vec{y} \in B(\vec{x}, \delta) \implies \|\vec{y} - \vec{a}\| \leq \|\vec{y} - \vec{x}\| + \|\vec{x} - \vec{a}\| < \delta + \|\vec{x} - \vec{a}\| = \epsilon.$$

This proves that  $B(\vec{a}, \epsilon)$  is open.

Since balls are open, by the second property in Proposition 6.4.4, the unions of balls are also open. Conversely, suppose  $U$  is open, then for any  $\vec{a} \in U$ , we have a ball  $B(\vec{a}, \epsilon_{\vec{a}}) \subset U$ , where the radius  $\epsilon_{\vec{a}}$  may depend on the point. Then

$$U \subset \cup_{\vec{a} \in U} \{\vec{a}\} \subset \cup_{\vec{a} \in U} B(\vec{a}, \epsilon_{\vec{a}}) \subset U$$

shows that  $U = \cup_{\vec{a} \in U} B(\vec{a}, \epsilon_{\vec{a}})$  is a union of balls.

To express  $U$  as a union of countably many balls, we use the fact that any vector has arbitrarily close rational vectors nearby. The fact can be directly verified for  $L^p$ -norms, and the general case will follow from Theorem 6.3.8. For each  $\vec{a} \in U$ , we have  $B(\vec{a}, \epsilon_{\vec{a}}) \subset U$ . By choosing smaller  $\epsilon_{\vec{a}}$  if necessary, we may further assume that  $\epsilon_{\vec{a}}$  is rational. Pick a rational vector  $\vec{r}_{\vec{a}} \in \mathbb{Q}^n$  satisfying  $d(\vec{r}_{\vec{a}}, \vec{a}) < \frac{\epsilon_{\vec{a}}}{2}$ . Then we have  $\vec{a} \in B\left(\vec{r}_{\vec{a}}, \frac{\epsilon_{\vec{a}}}{2}\right) \subset B(\vec{a}, \epsilon_{\vec{a}})$ . The same argument as before gives us

$$U = \cup_{\vec{a} \in U} B\left(\vec{r}_{\vec{a}}, \frac{\epsilon_{\vec{a}}}{2}\right).$$

Since the collection of all balls with rational center and rational radius is countable, the union on the right is actually countable (after deleting the duplicates).  $\square$

**Exercise 6.80.** For any subset  $A \subset \mathbb{R}^n$  and  $\epsilon > 0$ , prove that the  $\epsilon$ -neighborhood

$$A^\epsilon = \{\vec{x}: \|\vec{x} - \vec{a}\| < \epsilon \text{ for some } \vec{a} \in A\}$$

is open. Moreover, prove that  $\cap_{\epsilon > 0} A^\epsilon = A$  if and only if  $A$  is closed.

**Exercise 6.81.** For any subset  $A \subset \mathbb{R}^n$ , prove that the *interior* of  $A$

$$\mathring{A} = \{\vec{x}: B(\vec{x}, \epsilon) \subset A \text{ for some } \epsilon > 0\}$$

is open.

**Exercise 6.82.** Prove that the interior  $\mathring{A}$  is the biggest open subset contained in  $A$ , and  $\mathring{A} = \mathbb{R}^n - \overline{\mathbb{R}^n - A}$ .

## Open Subsets in $\mathbb{R}$

**Theorem 6.4.6.** *Any open subset of  $\mathbb{R}$  is a disjoint union of countably many open intervals. Moreover, the decomposition is unique.*

*Proof.* Let  $U \subset \mathbb{R}$  be open. If  $U = \sqcup (a_i, b_i)$ , then it is easy to see that  $(a_i, b_i)$  are maximal open intervals in the sense that if  $(a_i, b_i) \subset (a, b) \subset U$ , then  $(a_i, b_i) = (a, b)$ . This already explains the uniqueness of the decomposition.

Since the union of two intervals sharing a point is still an interval, it is easy to see that any two maximal open intervals must be either identical or disjoint. The (disjoint) union of maximal open intervals is contained in  $U$ . It remains to show that any point  $x \in U$  also lies in a maximal open interval. In fact, this maximal interval should be

$$(a_x, b_x) = \cup \{(a, b) : x \in (a, b) \subset U\},$$

where  $a_x = \inf\{a : x \in (a, b) \subset U\}$  and  $b$  is the similar supremum. First, the collection on the right is not empty because  $U$  is open. Therefore  $(a_x, b_x)$  is defined. Moreover, if  $(a_x, b_x) \subset (a, b) \subset U$ , then  $x \in (a, b)$ , so that  $(a, b)$  belongs to the collection on the right. Therefore  $(a, b) \subset (a_x, b_x)$ . This proves that  $(a_x, b_x)$  is maximal.

So we have proved that  $U = \sqcup (a_i, b_i)$ , where  $(a_i, b_i)$  are all the maximal open intervals in  $U$ . For any  $m \in \mathbb{N}$  and  $\epsilon > 0$ , by maximal open intervals being disjoint, the number of maximal open intervals in  $U$  that are contained in  $(-m, m)$  and have length  $b_i - a_i \geq \frac{1}{n}$  is  $\leq 2mn$ . Therefore the collection

$$\left\{ (a_i, b_i) : b_i - a_i \geq \frac{1}{n}, (a_i, b_i) \subset (-m, m) \right\}$$

is finite. Since any bounded  $(a_i, b_i)$  is contained in some  $(-m, m)$  and satisfies  $b_i - a_i \geq \frac{1}{n}$  for some  $n \in \mathbb{N}$ , we conclude that the countable collection

$$\cup_{m, n \in \mathbb{N}} \left\{ (a_i, b_i) : b_i - a_i \geq \frac{1}{n}, (a_i, b_i) \subset (-m, m) \right\}$$

is all the bounded maximal open intervals in  $U$ . The disjoint property implies that  $U$  has at most two unbounded maximal open intervals. Therefore the number of maximal open intervals in  $U$  is still countable.  $\square$

## Heine-Borel Theorem

**Theorem 6.4.7.** *Suppose  $K$  is a compact subset. Suppose  $\{U_i\}$  is a collection of open subsets such that  $K \subset \cup U_i$ . Then  $K \subset U_{i_1} \cup U_{i_2} \cup \cdots \cup U_{i_k}$  for finitely many open subsets in the collection.*

The single variable version of the theorem is given by Theorem 1.5.6. For the  $L^\infty$ -norm, the original proof can be easily adopted to the multivariable version. The compact subset  $K$  is contained in a bounded rectangle  $I = [\alpha_1, \beta_1] \times$

$[\alpha_2, \beta_2] \times \cdots \times [\alpha_n, \beta_n]$ . By replacing each interval  $[\alpha_i, \beta_i]$  with either  $\left[\alpha_i, \frac{\alpha_i + \beta_i}{2}\right]$  or  $\left[\frac{\alpha_i + \beta_i}{2}, \beta_i\right]$ , the rectangle can be divided into  $2^n$  rectangles. If  $K$  cannot be covered by finitely many open subsets from  $\{U_i\}$ , then for one of the  $2^n$  rectangles, denoted  $I_1$ , the intersection  $K_1 = K \cap I_1$  cannot be covered by finitely many open subsets from  $\{U_i\}$ . The rest of the construction and the argument can proceed as before.

The property in Theorem 6.4.7 is actually the official definition of compactness in general topological spaces. Therefore the property should imply our definition of compactness in terms of the sequence limit.

**Exercise 6.83.** Suppose  $K$  satisfies the property in Theorem 6.4.7. Use the open cover  $U_k = B(\vec{0}, k)$  to prove that  $K$  is bounded.

**Exercise 6.84.** Suppose  $\vec{x}_k \in K$  converges to  $\vec{a} \notin K$ . Use the open cover  $U_k = \mathbb{R}^n - \bar{B}(\vec{a}, k^{-1})$  (complements of closed balls around  $\vec{a}$ ) to prove that  $K$  does not have the property in Theorem 6.4.7.

**Exercise 6.85.** Prove the converse of Theorem 6.4.7.

## Set Theoretical Interpretation of Continuity

Let  $f: X \rightarrow Y$  be a map. Then the *image* of a subset  $A \subset X$  is

$$f(A) = \{f(x) : x \in A\} \subset Y,$$

and the *preimage* of a subset  $B \subset Y$  is

$$f^{-1}(B) = \{x \in X : f(x) \in B\} \subset X.$$

It is easy to verify that the following properties.

- $A \subset f^{-1}(B) \iff f(A) \subset B$ .
- $f^{-1}(f(A)) \subset A$ ,  $f(f^{-1}(B)) = B \cap f(X)$ .
- $A \subset A' \implies f(A) \subset f(A')$ .
- $B \subset B' \implies f^{-1}(B) \subset f^{-1}(B')$  and  $f^{-1}(B' - B) = f^{-1}(B') - f^{-1}(B)$ .
- $f^{-1}(Y - B) = X - f^{-1}(B)$ .
- $f(\cup A_i) = \cup f(A_i)$ ,  $f(\cap A_i) \subset \cap f(A_i)$ .
- $f^{-1}(\cup B_i) = \cup f^{-1}(B_i)$ ,  $f^{-1}(\cap B_i) = \cap f^{-1}(B_i)$ .
- $(g \circ f)(A) = g(f(A))$ ,  $(g \circ f)^{-1}(B) = f^{-1}(g^{-1}(B))$ .

**Proposition 6.4.8.** For a map  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the following are equivalent.

1.  $F$  is continuous.
2. The preimage  $F^{-1}(U)$  of any open  $U$  is open.
3. The preimage  $F^{-1}(B(\vec{b}, \epsilon))$  of any ball is open.
4. The preimage  $F^{-1}(C)$  of any closed  $C$  is closed.

If  $F$  is defined on a subset  $A \subset \mathbb{R}^n$ , then  $F^{-1}(B)$  is open (closed) means that  $F^{-1}(B) = A \cap D$  for a open (closed) subset  $D \subset \mathbb{R}^n$ .

The equivalence between the first and the fourth statements is already proved in Exercises 6.68 and 6.73, using sequence argument and with the help of Exercises 6.20 and 6.72.

*Proof.* Suppose  $F$  is continuous and  $U$  is open. Then for any  $\vec{a} \in F^{-1}(U)$ , we have  $F(\vec{a}) \in U$ . Since  $U$  is open, we have  $B(F(\vec{a}), \epsilon) \subset U$  for some  $\epsilon > 0$ . The continuity of  $F$  at  $\vec{a}$  means that, for this  $\epsilon$ , there is  $\delta > 0$ , such that

$$\|\vec{x} - \vec{a}\| < \delta \implies \|F(\vec{x}) - F(\vec{a})\| < \epsilon.$$

The implication is the same as

$$\vec{x} \in B(\vec{a}, \delta) \implies F(\vec{x}) \in B(F(\vec{a}), \epsilon).$$

Since  $B(F(\vec{a}), \epsilon) \subset U$ , this tells us  $F(B(\vec{a}, \delta)) \subset U$ , which is the same as  $B(\vec{a}, \delta) \subset F^{-1}(U)$ . Thus we have shown that

$$\vec{a} \in F^{-1}(U) \implies B(\vec{a}, \delta) \subset F^{-1}(U) \text{ for some } \delta > 0.$$

This proves that  $F^{-1}(U)$  is open. So the first statement implies the second.

By Proposition 6.4.5, any ball is open. Therefore the third statement is a special case of the second statement.

Now we prove the third statement implies the first. For any  $\vec{a}$  and  $\epsilon > 0$ , the third statement says that  $F^{-1}(B(F(\vec{a}), \epsilon))$  is open. On the other hand,  $F(\vec{a}) \in B(F(\vec{a}), \epsilon)$  implies  $\vec{a} \in F^{-1}(B(F(\vec{a}), \epsilon))$ . Therefore the openness implies  $B(\vec{a}, \delta) \subset F^{-1}(B(F(\vec{a}), \epsilon))$  for some  $\delta > 0$ . The meaning of the inclusion is exactly

$$\|\vec{x} - \vec{a}\| < \delta \implies \|F(\vec{x}) - F(\vec{a})\| < \epsilon.$$

This is the first statement.

It remains to show that the second and the fourth statements are equivalent. This follows from Proposition 6.4.3 and the property  $F^{-1}(\mathbb{R}^m - B) = \mathbb{R}^n - F^{-1}(B)$ . Suppose the second statement is true. Then for any closed  $B$ , the complement  $\mathbb{R}^m - B$  is open. The second statement tells us that the preimage  $F^{-1}(\mathbb{R}^m - B)$  is also open. Then  $F^{-1}(\mathbb{R}^m - B) = \mathbb{R}^n - F^{-1}(B)$  tells us that  $F^{-1}(B)$  is closed. The prove that the fourth statement implies the second is the same.  $\square$

**Exercise 6.86.** Prove the following are equivalent for a map  $F$  defined on the whole Euclidean space. For the definition of closure, see Exercise 6.70.

1. The map is continuous.
2.  $F(\bar{A}) \subset \overline{F(A)}$  for any subset  $A$ .
3.  $F^{-1}(\bar{B}) \supset \overline{F^{-1}(B)}$  for any subset  $B$ .

## 6.5 Additional Exercise

### Topology in a Subset of $\mathbb{R}^n$

The discussion about the open and closed subsets of  $\mathbb{R}^n$  can be extended to open and closed *relative to* a subset  $X$  of  $\mathbb{R}^n$ .

Exercise 6.87. We say a subset  $A \subset X$  is *closed in  $X$*  (or with respect to  $X$ ) if

$$\vec{x}_n \in A, \lim_{n \rightarrow \infty} \vec{x}_n = \vec{l} \in X \implies \vec{l} \in A.$$

Prove that this happens if and only if  $A = X \cap C$  for a closed subset  $C$  of  $\mathbb{R}^n$ .

Exercise 6.88. We say a subset  $A \subset X$  is *open in  $X$*  (or with respect to  $X$ ) if

$$\vec{x} \in A \implies B(\vec{x}, \epsilon) \cap X \subset A \text{ for some } \epsilon > 0.$$

In other words, for any point in  $A$ , all the points of  $X$  that are sufficiently near the point are also in  $A$ . Prove that this happens if and only if  $A = X \cap U$  for an open subset  $U$  of  $\mathbb{R}^n$ .

Exercise 6.89. For  $A \subset X \subset \mathbb{R}^n$ , prove that  $A$  is open in  $X$  if and only if  $X - A$  is closed in  $X$ .

Exercise 6.90. Extend Propositions 6.3.7 and 6.4.4 to subsets of  $X \subset \mathbb{R}^n$ .

### Repeated Extreme

Exercise 6.91. For a function  $f(\vec{x}, \vec{y})$  on  $A \times B$ . Prove that

$$\inf_{\vec{y} \in B} \sup_{\vec{x} \in A} f(\vec{x}, \vec{y}) \geq \sup_{\vec{x} \in A} \inf_{\vec{y} \in B} f(\vec{x}, \vec{y}) \geq \inf_{\vec{x} \in A} \inf_{\vec{y} \in B} f(\vec{x}, \vec{y}) = \inf_{\vec{x} \in A, \vec{y} \in B} f(\vec{x}, \vec{y}).$$

Exercise 6.92. For any  $a \geq b_1 \geq c_1 \geq d$  and  $a \geq b_2 \geq c_2 \geq d$ , can you construct a function  $f(x, y)$  on  $[0, 1] \times [0, 1]$  such that

$$\begin{array}{lll} \sup_{\vec{x} \in A, \vec{y} \in B} f = a, & \inf_{\vec{y} \in B} \sup_{\vec{x} \in A} f = b_1, & \inf_{\vec{x} \in A} \sup_{\vec{y} \in B} f = b_2, \\ \inf_{\vec{x} \in A, \vec{y} \in B} f = d, & \sup_{\vec{x} \in A} \inf_{\vec{y} \in B} f = c_1, & \sup_{\vec{y} \in B} \inf_{\vec{x} \in A} f = c_2. \end{array}$$

Exercise 6.93. Suppose  $f(x, y)$  is a function on  $[0, 1] \times [0, 1]$ . Prove that if  $f(x, y)$  is increasing in  $x$ , then  $\inf_{y \in [0, 1]} \sup_{x \in [0, 1]} f(x, y) = \sup_{x \in [0, 1]} \inf_{y \in [0, 1]} f(x, y)$ . Can the interval  $[0, 1]$  be changed to  $(0, 1)$ ?

Exercise 6.94. Extend the discussion to three or more repeated extrema. For example, can you give a simple criterion for comparing two strings of repeated extrema?

### Homogeneous and Weighted Homogeneous Function

A function  $f(\vec{x})$  is *homogeneous* of degree  $p$  if  $f(c\vec{x}) = c^p f(\vec{x})$  for any  $c > 0$ . More generally, a function is *weighted homogeneous* if

$$cf(x_1, x_2, \dots, x_n) = f(c^{q_1} x_1, c^{q_2} x_2, \dots, c^{q_n} x_n).$$

The later part of the proof of Theorem 6.3.8 makes use of the fact that any norm is a homogeneous function of degree 1 (also see Exercise 6.97).

**Exercise 6.95.** Prove that two homogeneous functions of the same degree are equal away from  $\vec{0}$  if and only if their restrictions on the unit sphere  $S^{n-1}$  are equal.

**Exercise 6.96.** Prove that a homogeneous function is bigger than another homogeneous function of the same degree away from  $\vec{0}$  if and only if the inequality holds on  $S^{n-1}$ .

**Exercise 6.97.** Suppose  $f(\vec{x})$  is a continuous homogeneous function of degree  $p$  satisfying  $f(\vec{x}) > 0$  for  $\vec{x} \neq \vec{0}$ . Prove that there is  $c > 0$ , such that  $f(\vec{x}) \geq c\|\vec{x}\|^p$  for any  $\vec{x}$ .

**Exercise 6.98.** Prove that a homogeneous function is continuous away from  $\vec{0}$  if and only if its restriction on  $S^{n-1}$  is continuous. Then find the condition for a homogeneous function to be continuous at  $\vec{0}$ .

### Continuous Map and Function on Compact Set

**Exercise 6.99.** Suppose  $F(\vec{x}, \vec{y})$  is a continuous map on  $A \times K \subset \mathbb{R}^m \times \mathbb{R}^n$ . Prove that if  $K$  is compact, then for any  $\vec{a} \in A$  and  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\vec{x} \in A$ ,  $\vec{y} \in K$ , and  $\|\vec{x} - \vec{a}\| < \delta$  imply  $\|F(\vec{x}, \vec{y}) - F(\vec{a}, \vec{y})\| < \epsilon$ .

**Exercise 6.100.** Suppose  $f(\vec{x}, \vec{y})$  is a continuous function on  $A \times K$ . Prove that if  $K$  is compact, then  $g(\vec{x}) = \max_{\vec{y} \in K} f(\vec{x}, \vec{y})$  is a continuous function on  $A$ .

**Exercise 6.101.** Suppose  $f(\vec{x}, y)$  is a continuous function on  $A \times [0, 1]$ . Prove that  $g(\vec{x}) = \int_0^1 f(\vec{x}, y) dy$  is a continuous function on  $A$ .

### Continuity in Coordinates

A function  $f(\vec{x}, \vec{y})$  is continuous in  $\vec{x}$  if  $\lim_{\vec{x} \rightarrow \vec{a}} f(\vec{x}, \vec{y}) = f(\vec{a}, \vec{y})$ . It is uniformly continuous in  $\vec{x}$  if the limit is uniform in  $\vec{y}$ : For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|\vec{x} - \vec{x}'\| < \delta$  implies  $\|f(\vec{x}, \vec{y}) - f(\vec{x}', \vec{y})\| \leq \epsilon$  for any  $\vec{y}$ .

**Exercise 6.102.** Prove that if  $f(\vec{x}, \vec{y})$  is continuous, then  $f(\vec{x}, \vec{y})$  is continuous in  $\vec{x}$ . Moreover, show that the function

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0), \end{cases}$$

is continuous in both  $x$  and  $y$ , but is not a continuous function.

**Exercise 6.103.** Prove that if  $f(\vec{x}, \vec{y})$  is continuous in  $\vec{x}$  and is uniformly continuous in  $\vec{y}$ , then  $f(\vec{x}, \vec{y})$  is continuous.

**Exercise 6.104.** Suppose  $f(\vec{x}, y)$  is continuous in  $\vec{x}$  and in  $y$ . Prove that if  $f(\vec{x}, y)$  is monotone in  $y$ , then  $f(\vec{x}, y)$  is continuous.





## Chapter 7

# Multivariable Algebra

## 7.1 Linear Transform

A *vector space* is a set in which addition  $\vec{v} + \vec{w}$  and scalar multiplication  $c\vec{v}$  are defined and satisfy the usual properties. A map  $L: V \rightarrow W$  between vector spaces is a *linear transform* if it preserves the two operations

$$L(\vec{x} + \vec{y}) = L(\vec{x}) + L(\vec{y}), \quad L(c\vec{x}) = cL(\vec{x}).$$

This implies that the *linear combinations* are also preserved

$$L(c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_k\vec{x}_k) = c_1L(\vec{x}_1) + c_2L(\vec{x}_2) + \cdots + c_kL(\vec{x}_k).$$

An *isomorphism* of vector spaces is an invertible linear transform.

**Exercise 7.1.** Suppose  $K, L: V \rightarrow W$  are linear transforms and  $c$  is a number. Prove that the maps  $(K + L)(\vec{x}) = K(\vec{x}) + L(\vec{x})$  and  $(cL)(\vec{x}) = c(L(\vec{x}))$  are linear transforms.

**Exercise 7.2.** Suppose  $L: U \rightarrow V$  and  $K: V \rightarrow W$  are linear transforms. Prove that the composition  $K \circ L: U \rightarrow W$  is a linear transform.

**Exercise 7.3.** Suppose  $L: V \rightarrow W$  is a continuous map satisfying  $L(\vec{x} + \vec{y}) = L(\vec{x}) + L(\vec{y})$ . Prove that the maps  $L$  is a linear transform.

Vector space is the natural generalization of the Euclidean space  $\mathbb{R}^n$ . In fact, a vector space will always be finite dimensional in this book, and any finite dimensional vector space is isomorphic to a Euclidean space.

The linear functions used in the differentiation of single variable functions will become maps of the form  $\vec{a} + L(\Delta\vec{x})$  in the differentiation of multivariable functions. We will call such maps *linear maps* (more properly named *affine maps*) to distinguish from linear transforms (for which  $\vec{a} = \vec{0}$ ).

## Matrix of Linear Transform

An ordered collection of vectors  $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  is a *basis* of  $V$  if any vector has unique expression

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n.$$

The coefficient  $x_i$  is the  $i$ -th *coordinate* of  $\vec{x}$  with respect to the basis. Suppose  $\alpha$  and  $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\}$  are ordered basis of  $V$  and  $W$ . Then a linear transform  $L: V \rightarrow W$  is determined by its values on the basis vectors

$$\begin{aligned} L(\vec{v}_1) &= a_{11}\vec{w}_1 + a_{21}\vec{w}_2 + \cdots + a_{m1}\vec{w}_m, \\ L(\vec{v}_2) &= a_{12}\vec{w}_1 + a_{22}\vec{w}_2 + \cdots + a_{m2}\vec{w}_m, \\ &\vdots \\ L(\vec{v}_n) &= a_{1n}\vec{w}_1 + a_{2n}\vec{w}_2 + \cdots + a_{mn}\vec{w}_m. \end{aligned}$$

The matrix of  $L$  with respect to the (ordered) bases  $\alpha$  and  $\beta$  is the transpose of the coefficients (also see Exercise 7.4)

$$L_{\beta\alpha} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The relation between a linear transform and its matrix can be expressed as

$$L\alpha = L(\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_n) = (\vec{w}_1 \ \vec{w}_2 \ \cdots \ \vec{w}_m)L_{\beta\alpha} = \beta L_{\beta\alpha}.$$

Here the left means applying  $L$  to each vector, and the right means the “matrix product” with  $\vec{w}_i$  viewed as columns.

The *standard basis*  $\epsilon = \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$  for the Euclidean space  $\mathbb{R}^n$  is given by

$$\vec{e}_1 = (1, 0, \dots, 0), \ \vec{e}_2 = (0, 1, \dots, 0), \ \dots, \ \vec{e}_n = (0, 0, \dots, 1).$$

The standard matrix for a linear transform  $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is (the two  $\epsilon$  may be different)

$$L_{\epsilon\epsilon} = (L(\vec{e}_1) \ L(\vec{e}_2) \ \cdots \ L(\vec{e}_n)).$$

Here the vectors  $L(\vec{e}_1) \in \mathbb{R}^m$  are written vertically and become the columns of the matrix. We may define the addition, scalar multiplication, and product of matrices as corresponding to the addition, scalar multiplication, and composition of linear transforms between Euclidean spaces.

**Exercise 7.4.** Suppose  $L: V \rightarrow W$  is a linear transform. Suppose  $\alpha$  and  $\beta$  are bases of  $V$  and  $W$ , and

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n, \quad L(\vec{x}) = y_1\vec{w}_1 + y_2\vec{w}_2 + \cdots + y_n\vec{w}_n.$$

Prove that

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

where the matrix is the matrix  $L_{\beta\alpha}$  of the linear transform.

**Exercise 7.5.** Suppose  $K, L: V \rightarrow W$  are linear transforms and  $c$  is a number. For the linear transforms  $K + L$  and  $cK$  in Exercise 7.1, show that  $(K + L)_{\beta\alpha} = K_{\beta\alpha} + L_{\beta\alpha}$ ,  $(cK)_{\beta\alpha} = cK_{\beta\alpha}$ .

**Exercise 7.6.** Suppose  $L: U \rightarrow V$  and  $K: V \rightarrow W$  are linear transforms. For the linear transform  $K \circ L$  in Exercise 7.2, show that  $(K \circ L)_{\gamma\alpha} = K_{\gamma\beta}L_{\beta\alpha}$ .

**Exercise 7.7.** Suppose  $L: V \rightarrow W$  is an invertible linear transform. Prove that  $L^{-1}: W \rightarrow V$  is also a linear transform, and  $(L^{-1})_{\alpha\beta} = (L_{\beta\alpha})^{-1}$ .

**Exercise 7.8.** Prove that the sum and scalar multiplication of linear transforms from  $V$  to  $W$  defined in Exercise 7.5 make the collection  $\text{Hom}(V, W)$  of all linear transforms from  $V$  to  $W$  into a vector space. Then prove the following.

1. Prove that  $(c_1K_1 + c_2K_2) \circ L = c_1K_1 \circ L + c_2K_2 \circ L$ . Then explain that for a linear transform  $L: U \rightarrow V$ , the map

$$\circ L: \text{Hom}(V, W) \rightarrow \text{Hom}(U, W).$$

is a linear transform.

2. For a linear transform  $K: V \rightarrow W$ , prove that the map

$$K \circ: \text{Hom}(U, V) \rightarrow \text{Hom}(U, W).$$

is a linear transform.

**Exercise 7.9.** Prove that a linear transform  $L$  is injective if and only if  $K \circ L = I$  for some linear transform  $K$ . Moreover,  $L$  is surjective if and only if  $L \circ K = I$  for some linear transform  $K$ .

## Change of Basis

Let  $\alpha$  and  $\beta$  be two bases of  $V$ . We can express the first basis in terms of the second bases

$$\begin{aligned}\vec{v}_1 &= p_{11}\vec{w}_1 + p_{21}\vec{w}_2 + \cdots + p_{n1}\vec{w}_n, \\ \vec{v}_2 &= p_{12}\vec{w}_1 + p_{22}\vec{w}_2 + \cdots + p_{n2}\vec{w}_n, \\ &\vdots \\ \vec{v}_n &= p_{1n}\vec{w}_1 + p_{2n}\vec{w}_2 + \cdots + p_{nn}\vec{w}_n.\end{aligned}$$

By viewing the left side as the identity linear transform applied to the first basis, the coefficient matrix

$$I_{\beta\alpha} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

is the matrix of the identity transform  $I: V \rightarrow V$  with respect to the bases  $\alpha$  (of first  $V$ ) and  $\beta$  (of second  $V$ ). This is the usual matrix for changing the basis from  $\alpha$  to  $\beta$ . The matrix is also characterized by

$$\alpha = \beta I_{\beta\alpha}.$$

Suppose  $L: V \rightarrow W$  is a linear transform. Suppose  $\alpha, \alpha'$  are bases of  $V$ , and  $\beta, \beta'$  are bases of  $W$ . Then we have

$$L\alpha' = L\alpha I_{\alpha\alpha'} = \beta L_{\beta\alpha} I_{\alpha\alpha'} = \beta' I_{\beta'\beta} L_{\beta\alpha} I_{\alpha\alpha'}.$$

This gives the *change of basis formula*

$$L_{\beta'\alpha'} = I_{\beta'\beta} L_{\beta\alpha} I_{\alpha\alpha'}.$$

**Exercise 7.10.** A linear transform  $L: V \rightarrow V$  from a vector space to itself is an *endomorphism*. We usually take  $\alpha = \beta$  in this case and say that  $L_{\alpha\alpha}$  is the matrix of  $L$  with respect to  $\alpha$ . Prove that the matrices  $A$  and  $B$  of an endomorphism  $L: V \rightarrow V$  with respect to different bases of  $V$  are similar in the sense that  $B = PAP^{-1}$  for an invertible  $P$ .

## Dual Space

A *linear functional* on a vector space  $V$  is a numerical valued linear transform  $l: V \rightarrow \mathbb{R}$ . Since the addition and scalar multiplication of linear functionals are still linear (see Exercises 7.5 and 7.8), all the linear functionals on  $V$  form a vector space  $V^*$ , called the *dual space* of  $V$ . A linear transform  $L: V \rightarrow W$  induces the *dual transform* in the opposite direction by simply sending  $l \in W^*$  to  $l \circ L \in V^*$

$$L^*: V^* \leftarrow W^*, \quad L^*(l)(\vec{x}) = l(L(\vec{x})).$$

Given a basis  $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  of  $V$ , we may construct a *dual basis*  $\alpha^* = \{\vec{v}_1^*, \vec{v}_2^*, \dots, \vec{v}_n^*\}$  of  $V^*$  by taking  $\vec{v}_i^*$  to be the  $i$ -th coordinate with respect to the basis  $\alpha$

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \dots + x_n\vec{v}_n \mapsto \vec{v}_i^*(\vec{x}) = x_i.$$

In other words, the dual basis is characterized by

$$\vec{x} = \vec{v}_1^*(\vec{x})\vec{v}_1 + \vec{v}_2^*(\vec{x})\vec{v}_2 + \dots + \vec{v}_n^*(\vec{x})\vec{v}_n.$$

Moreover,  $\alpha^*$  is indeed a basis of  $V^*$  because any linear functional  $l$  has unique linear expression

$$l = a_1\vec{v}_1^* + a_2\vec{v}_2^* + \dots + a_n\vec{v}_n^*, \quad a_i = l(\vec{v}_i),$$

or

$$l = l(\vec{v}_1)\vec{v}_1^* + l(\vec{v}_2)\vec{v}_2^* + \dots + l(\vec{v}_n)\vec{v}_n^*.$$

**Proposition 7.1.1.** Suppose  $V$  is a finite dimensional vector space. Then

$$\vec{x} \mapsto \vec{x}^{**}, \quad \vec{x}^{**}(l) = l(\vec{x})$$

is an isomorphism  $V \cong (V^*)^*$ .

*Proof.* The following shows that  $\vec{x}^{**}$  is indeed a linear functional on  $V^*$  and therefore belongs to the double dual  $(V^*)^*$

$$\vec{x}^{**}(c_1l_1 + c_2l_2) = (c_1l_1 + c_2l_2)(\vec{x}) = c_1l_1(\vec{x}) + c_2l_2(\vec{x}) = c_1\vec{x}^{**}(l_1) + c_2\vec{x}^{**}(l_2).$$

The following shows that  $\vec{x}^{**}$  is linear

$$\begin{aligned} (c_1\vec{x}_1 + c_2\vec{x}_2)^{**}(l) &= l(c_1\vec{x}_1 + c_2\vec{x}_2) = c_1l(\vec{x}_1) + c_2l(\vec{x}_2) \\ &= c_1\vec{x}_1^{**}(l) + c_2\vec{x}_2^{**}(l) = (c_1\vec{x}_1^{**} + c_2\vec{x}_2^{**})(l). \end{aligned}$$

The following shows that  $\vec{x} \mapsto \vec{x}^{**}$  is injective

$$\vec{x}^{**} = 0 \implies l(\vec{x}) = \vec{x}^{**}(l) = 0 \text{ for all } l \in V^* \implies \vec{x} = \vec{0}.$$

In the second implication, if  $\vec{x} \neq \vec{0}$ , then we can expand  $\vec{x}$  to a basis and then construct a linear functional  $l$  such that  $l(\vec{x}) \neq 0$ .

Finally, the discussion about dual basis tells us that  $V^*$  and  $V$  have the same dimension. Then  $(V^*)^*$  and  $V^*$  also have the same dimension. Therefore  $(V^*)^*$  and  $V$  have the same dimension. The injective linear transform  $\vec{x} \mapsto \vec{x}^{**}$  between vector spaces of the same dimension must be an isomorphism.  $\square$

**Exercise 7.11.** Prove that  $(V \oplus W)^* = V^* \oplus W^*$ .

**Exercise 7.12.** Prove that the dual transform  $L^*$  is linear. This is a special case of the second part of Exercise 7.8.

**Exercise 7.13.** Prove that  $(K + L)^* = K^* + L^*$ ,  $(cL)^* = cL^*$ ,  $(K \circ L)^* = L^* \circ K^*$ .

**Exercise 7.14.** Prove that  $(L^*)^*$  corresponds to  $L$  under the isomorphism given by Proposition 7.1.1.

**Exercise 7.15.** Prove that  $L$  is injective if and only if  $L^*$  is surjective, and  $L$  is surjective if and only if  $L^*$  is injective.

**Exercise 7.16.** Prove that  $(L^*)_{\alpha^* \beta^*} = (L_{\beta \alpha})^T$ . This means that if  $A$  is the matrix of  $L$  with respect to some bases, then  $A^T$  is the matrix of  $L^*$  with respect to the dual bases. We also note the special case  $I_{\alpha^* \beta^*} = (I_{\beta \alpha})^T$  for the base change matrix.

## Norm of Linear Transform

Given a linear transform  $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$  and a norm on  $\mathbb{R}^m$ , we have

$$\begin{aligned} \|L(\vec{x})\| &\leq |x_1| \|\vec{a}_1\| + |x_2| \|\vec{a}_2\| + \cdots + |x_n| \|\vec{a}_n\| \\ &\leq (\|\vec{a}_1\| + \|\vec{a}_2\| + \cdots + \|\vec{a}_n\|) \|\vec{x}\|_\infty. \end{aligned}$$

By isomorphisms between general vector spaces and Euclidean spaces, and by the equivalence of norms (Theorem 6.3.8), for any linear transform  $L: V \rightarrow W$  of normed vector spaces, we can find  $\lambda$ , such that  $\|L(\vec{x})\| \leq \lambda \|\vec{x}\|$ . Then by

$$\frac{\|L(c\vec{x})\|}{\|c\vec{x}\|} = \frac{\|cL(\vec{x})\|}{\|c\vec{x}\|} = \frac{|c| \|L(\vec{x})\|}{|c| \|\vec{x}\|} = \frac{\|L(\vec{x})\|}{\|\vec{x}\|},$$

the smallest such  $\lambda$  is

$$\begin{aligned} \|L\| &= \inf \left\{ \lambda : \lambda \geq \frac{\|L(\vec{x})\|}{\|\vec{x}\|} \text{ for all } \vec{x} \right\} \\ &= \sup \left\{ \frac{\|L(\vec{x})\|}{\|\vec{x}\|} : \text{all } \vec{x} \in V \right\} \\ &= \sup \{ \|L(\vec{x})\| : \|\vec{x}\| = 1 \}. \end{aligned}$$

The number  $\|L\|$  is the *norm of the linear transform* with respect to the given norms. The inequality

$$\|L(\vec{x}) - L(\vec{y})\| = \|L(\vec{x} - \vec{y})\| \leq \|L\| \|\vec{x} - \vec{y}\|$$

also implies that linear transforms from finitely dimensional vector spaces are continuous.

**Proposition 7.1.2.** *The norm of linear transform satisfies the three axioms for the norm. Moreover, the norm of composition satisfies  $\|K \circ L\| \leq \|K\| \|L\|$ .*

By Exercise 7.8, the collection  $\text{Hom}(V, W)$  of all linear transforms from  $V$  to  $W$  is a vector space. The proposition shows that the norm of linear transforms is indeed a norm of this vector space.

*Proof.* By the definition, we have  $\|L\| \geq 0$ . If  $\|L\| = 0$ , then  $\|L(\vec{x})\| = 0$  for any  $\vec{x}$ . By the positivity of norm, we get  $L(\vec{x}) = \vec{0}$  for any  $\vec{x}$ . In other words,  $L$  is the zero transform. This verifies the positivity of  $\|L\|$ .

By the definition of the norm of linear transform, we have

$$\begin{aligned} \|K(\vec{x}) + L(\vec{x})\| &\leq \|K(\vec{x})\| + \|L(\vec{x})\| \leq \|K\| \|\vec{x}\| + \|L\| \|\vec{x}\| \leq (\|K\| + \|L\|) \|\vec{x}\|, \\ \|(cL)(\vec{x})\| &= \|cL(\vec{x})\| = |c| \|L(\vec{x})\| \leq |c| \|L\| \|\vec{x}\|, \\ \|(K \circ L)(\vec{x})\| &= \|K(L(\vec{x}))\| \leq \|K\| \|L(\vec{x})\| \leq \|K\| \|L\| \|\vec{x}\|. \end{aligned}$$

This gives  $\|K + L\| \leq \|K\| + \|L\|$ ,  $\|cL\| = |c| \|L\|$  and  $\|K \circ L\| \leq \|K\| \|L\|$ .  $\square$

**Example 7.1.1.** We want to find the norm of a linear functional  $l(\vec{x}) = \vec{a} \cdot \vec{x}: \mathbb{R}^n \rightarrow \mathbb{R}$  with respect to the Euclidean norm on  $\mathbb{R}^n$  and the absolute value on  $\mathbb{R}$ . Schwarz's inequality tells us  $|l(\vec{x})| \leq \|\vec{a}\|_2 \|\vec{x}\|_2$ . On the other hand, by  $\vec{a} \cdot \vec{a} = \|\vec{a}\|_2^2$ , the equality holds if we take  $\vec{x} = \vec{a}$ . Therefore the norm of the linear functional is  $\|l\| = \|\vec{a}\|_2$ .

**Example 7.1.2.** By using matrices, the collection of all the linear transforms from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  is identified with the Euclidean space  $\mathbb{R}^{mn}$ . On the space  $\mathbb{R}^{mn}$ , however, the preferred norm is the norm of linear transform.

A  $2 \times 2$  matrix with coefficients  $(a, b, c, d) \in \mathbb{R}^4$  gives a linear transform  $L(x, y) = (ax + by, cx + dy)$  of  $\mathbb{R}^2$  to itself. With respect to the  $L^\infty$ -norm of  $\mathbb{R}^2$ , we have

$$\begin{aligned} \|L(x, y)\|_\infty &= \max\{|ax + by|, |cx + dy|\} \\ &\leq \max\{|a| + |b|, |c| + |d|\} \max\{|x|, |y|\} = \max\{|a| + |b|, |c| + |d|\} \|(x, y)\|_\infty. \end{aligned}$$

Moreover, the equality is satisfied for  $x = \pm 1$  and  $y = \pm 1$  for suitable choice of the signs. Therefore  $\|L\| = \max\{|a| + |b|, |c| + |d|\}$ . We have

$$\frac{1}{2} \|(a, b, c, d)\|_1 \leq \|L\| \leq \|(a, b, c, d)\|_1.$$

**Example 7.1.3.** The linear transforms from a vector space  $V$  to itself form a vector space  $\text{End}(V) = \text{Hom}(V, V)$ . For a norm on  $V$ , we have the norm of linear transform on  $\text{End}(V)$  with respect to the norm on  $V$ . The square map  $F(L) = L^2: \text{End}(V) \rightarrow \text{End}(V)$  is

basically taking the square of square matrices. If  $H \in \text{End}(V)$  satisfies  $\|H\| < \epsilon$ , then by Proposition 7.1.2, we have

$$\begin{aligned} \|(L + H)^2 - L^2\| &= \|LH + HL + H^2\| \leq \|LH\| + \|HL\| + \|H^2\| \\ &\leq \|L\|\|H\| + \|H\|\|L\| + \|H\|^2 < 2\epsilon\|L\| + \epsilon^2. \end{aligned}$$

It is then easy to deduce the continuity of the square map.

**Exercise 7.17.** Extend Examples 7.1.1 and 7.1.2 to linear functionals on an inner product space.

**Exercise 7.18.** Find the norm of a linear transform  $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , where  $\mathbb{R}^n$  has the Euclidean norm and  $\mathbb{R}^m$  has the  $L^\infty$ -norm.

**Exercise 7.19.** Extend Example 7.1.2 to linear transforms on other Euclidean spaces and other norms.

**Exercise 7.20.** Suppose  $U, V, W$  are normed vector spaces. Prove that the linear transforms  $\circ L$  and  $K \circ$  in Exercise 7.8 satisfy  $\|\circ L\| \leq \|L\|$  and  $\|K \circ\| \leq \|K\|$ .

**Exercise 7.21.** Consider the normed vector space  $\text{End}(V)$  in Example 7.1.3. Let  $GL(V) \subset \text{End}(V)$  be the subset of invertible linear transforms.

1. Suppose  $\|L\| < 1$ . Prove that  $\sum_{n=0}^{\infty} L^n = I + L + L^2 + \cdots$  converges in  $\text{End}(V)$  and is the inverse of  $I - L$ .
2. Suppose  $L$  is invertible and  $\|K - L\| < \frac{1}{\|L^{-1}\|}$ . By taking  $I - KL^{-1}$  to be  $L$  in the first part, prove that  $K$  is also invertible.
3. In the second part, further prove that

$$\|K^{-1}\| \leq \frac{\|L^{-1}\|}{1 - \|L - K\|\|L^{-1}\|}, \quad \|K^{-1} - L^{-1}\| \leq \frac{\|K - L\|\|L^{-1}\|^2}{1 - \|K - L\|\|L^{-1}\|}.$$

4. Prove that  $GL(V)$  is an open subset of  $\text{End}(V)$ , and the inverse  $L \mapsto L^{-1}$  is a continuous map on  $GL(V)$ .

**Exercise 7.22.** Prove that the supremum in the definition of the norm of linear transform is in fact the maximum. In other words, the supremum is reached at some vector of unit length.

**Exercise 7.23.** Prove that with respect to the norm of linear transform, the addition, scalar multiplication and composition of linear transforms are continuous maps.

## 7.2 Bilinear Map

Let  $U, V, W$  be vector spaces. A map  $B: V \times W \rightarrow U$  is *bilinear* if it is linear in both vector variables

$$\begin{aligned} B(\vec{x} + \vec{x}', \vec{y}) &= B(\vec{x}, \vec{y}) + B(\vec{x}', \vec{y}), & B(c\vec{x}, \vec{y}) &= cB(\vec{x}, \vec{y}), \\ B(\vec{x}, \vec{y} + \vec{y}') &= B(\vec{x}, \vec{y}) + B(\vec{x}, \vec{y}'), & B(\vec{x}, c\vec{y}) &= cB(\vec{x}, \vec{y}). \end{aligned}$$



The scalar product and dot product (and more generally inner product)

$$c\vec{x}: \mathbb{R} \times V \rightarrow V, \quad \vec{x} \cdot \vec{y}: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$$

are bilinear maps. The 3-dimensional *cross product*

$$(x_1, x_2, x_3) \times (y_1, y_2, y_3) = (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1): \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

and the matrix product (or composition of linear transforms)

$$AB: \mathbb{R}^{mn} \times \mathbb{R}^{nk} \rightarrow \mathbb{R}^{mk}$$

are also bilinear. Therefore bilinear maps can be considered as generalized products.

The addition and scalar multiplication of bilinear maps are still bilinear. Moreover, if  $B$  is bilinear and  $K$  and  $L$  are linear, then  $B(K(\vec{x}), L(\vec{y}))$  and  $L(B(\vec{x}, \vec{y}))$  are bilinear.

Let  $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  and  $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\}$  be bases of  $V$  and  $W$ . Let  $x_i$  and  $y_j$  be the coordinates of  $\vec{x} \in V$  and  $\vec{y} \in W$  with respect to the bases. Then

$$B(\vec{x}, \vec{y}) = B\left(\sum_i x_i \vec{v}_i, \sum_j y_j \vec{w}_j\right) = \sum_{i,j} x_i y_j \vec{b}_{ij}, \quad \vec{b}_{ij} = B(\vec{v}_i, \vec{w}_j).$$

If  $U, V, W$  are Euclidean spaces, and  $\alpha$  and  $\beta$  are the standard bases, then

$$\|B(\vec{x}, \vec{y})\| \leq \sum_{i,j} |x_i| |y_j| \|\vec{b}_{ij}\| \leq \left( \sum_{i,j} \|\vec{b}_{ij}\| \right) \|\vec{x}\|_\infty \|\vec{y}\|_\infty.$$

For finite dimensional normed vector spaces  $U, V$  and  $W$ , by isomorphisms between general vector spaces and Euclidean spaces, and by the equivalence of norms, we get  $\|B(\vec{x}, \vec{y})\| \leq \lambda \|\vec{x}\| \|\vec{y}\|$  for a constant  $\lambda$ . This implies that the bilinear map is continuous, and we may define the smallest such  $\lambda$  as the *norm of the bilinear map*

$$\begin{aligned} \|B\| &= \inf\{\lambda: \|B(\vec{x}, \vec{y})\| \leq \lambda \|\vec{x}\| \|\vec{y}\| \text{ for all } \vec{x}, \vec{y}\} \\ &= \sup\{\|B(\vec{x}, \vec{y})\|: \|\vec{x}\| = \|\vec{y}\| = 1\}. \end{aligned}$$

**Exercise 7.24.** Prove that any bilinear map from finite dimensional vector spaces is continuous.

**Exercise 7.25.** Find the norms of the scalar product and dot product with respect to the Euclidean norm.

**Exercise 7.26.** Prove that the norm of bilinear map satisfies the three axioms for the norm.

**Exercise 7.27.** How is the norm of the bilinear map  $B(K(\vec{x}), L(\vec{y}))$  related to the norms of the bilinear map  $B$  and linear transforms  $K$  and  $L$ ?

## Bilinear Function

A *bilinear function* is a numerical valued bilinear map  $b: V \times W \rightarrow \mathbb{R}$ . Given bases  $\alpha$  and  $\beta$  of  $V$  and  $W$ , the bilinear function is

$$b(\vec{x}, \vec{y}) = \sum_{i,j} b_{ij} x_i y_j, \quad b_{ij} = b(\vec{v}_i, \vec{w}_j).$$

This gives a one-to-one correspondence between bilinear functions and matrices  $B_{\alpha\beta} = (b_{ij})$ .

A bilinear function induces two linear transforms

$$V \rightarrow W^*: \vec{x} \mapsto b(\vec{x}, \cdot), \quad W \rightarrow V^*: \vec{y} \mapsto b(\cdot, \vec{y}). \quad (7.2.1)$$

Exercise 7.29 shows that the two linear transforms determine each other. Conversely, a linear transform  $L: V \rightarrow W^*$  gives a bilinear form  $b(\vec{x}, \vec{y}) = L(\vec{x})(\vec{y})$ . Here  $L(\vec{x})$  is a linear functional on  $W$  and is applied to a vector  $\vec{y}$  in  $W$ . Therefore bilinear functions on  $V \times W$  may also be identified with  $\text{Hom}(V, W^*)$ .

**Exercise 7.28.** How is the matrix  $B_{\alpha\beta}$  of a bilinear function changed if the bases are changed?

**Exercise 7.29.** Prove that the two linear transforms (7.2.1) are dual to each other, after applying Proposition 7.1.1.

**Exercise 7.30.** Let  $\alpha$  and  $\beta$  be bases of  $V$  and  $W$ . Let  $\alpha^*$  and  $\beta^*$  be dual bases of  $V^*$  and  $W^*$ . Prove that the matrix of the induced linear transform  $V \rightarrow W^*$  with respect to the bases  $\alpha$  and  $\beta^*$  is  $B_{\alpha\beta}$ . What is the matrix of the induced linear transform  $W \rightarrow V^*$  with respect to the bases  $\beta$  and  $\alpha^*$ ?

**Exercise 7.31.** Prove that bilinear functions  $b$  on  $\mathbb{R}^m \times \mathbb{R}^n$  are in one-to-one correspondence with linear transforms  $L: \mathbb{R}^m \rightarrow \mathbb{R}^n$  by

$$b(\vec{x}, \vec{y}) = L(\vec{x}) \cdot \vec{y}.$$

1. What is the relation between the matrix of  $L$  and the matrix of  $b$  with respect to the standard bases?
2. Prove that the norms of  $b$  and  $L$  with respect to the Euclidean norms are equal. In other words, we have

$$\|L\| = \sup_{\|\vec{x}\|=\|\vec{y}\|=1} L(\vec{x}) \cdot \vec{y}.$$

3. Extend the discussion to bilinear functions on inner product spaces.

## Dual Pairing

A bilinear function is a *dual pairing* if both induced linear transforms (7.2.1) are injective. The injections imply

$$\dim V \leq \dim W^* = \dim W, \quad \dim W \leq \dim V^* = \dim V,$$

so that  $\dim V = \dim W$ , and both linear transforms are isomorphisms. Therefore a dual pairing identifies  $W$  with the dual space  $V^*$  and identifies  $V$  with the dual space  $W^*$ . In particular, for a basis  $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  of  $V$ , we may define the dual basis  $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$  of  $W$  with respect to the dual pairing  $b$  as the basis corresponding to the dual basis  $\alpha^*$  of  $V^*$  under the isomorphism  $W \cong V^*$ . This means that the dual basis is characterized by

$$b(\vec{v}_i, \vec{w}_j) = \delta_{ij}.$$

This implies the expression of any  $\vec{v} \in V$  in terms of  $\alpha$

$$\vec{v} = b(\vec{v}, \vec{w}_1)\vec{v}_1 + b(\vec{v}, \vec{w}_2)\vec{v}_2 + \dots + b(\vec{v}, \vec{w}_n)\vec{v}_n, \quad (7.2.2)$$

and the similar expression of vectors in  $W$  in terms of  $\beta$ .

**Proposition 7.2.1.** *Suppose  $b: V \times W \rightarrow \mathbb{R}$  is a dual pairing. Then  $\lim \vec{v}_n = \vec{v}$  in  $V$  if and only if  $\lim b(\vec{v}_n, \vec{w}) = b(\vec{v}, \vec{w})$  for all  $\vec{w} \in W$ , and  $\lim \vec{w}_n = \vec{w}$  in  $W$  if and only if  $\lim b(\vec{v}, \vec{w}_n) = b(\vec{v}, \vec{w})$  for all  $\vec{v} \in V$ .*

*Proof.* Since bilinear forms are continuous, we get  $\lim \vec{v}_n = \vec{v}$  implying  $\lim b(\vec{v}_n, \vec{w}) = b(\vec{v}, \vec{w})$  for all  $\vec{w} \in W$ . Conversely, if  $\lim b(\vec{v}_n, \vec{w}) = b(\vec{v}, \vec{w})$  for all  $\vec{w} \in W$ , then  $\lim b(\vec{v}_n, \vec{w}_i) = b(\vec{v}, \vec{w}_i)$  for  $i = 1, 2, \dots, n$ , and it further follows from (7.2.2) that  $\lim \vec{v}_n = \vec{v}$ .  $\square$

**Example 7.2.1.** The *evaluation pairing* is

$$\langle \vec{x}, l \rangle = l(\vec{x}): V \times V^* \rightarrow \mathbb{R}.$$

The first isomorphism in (7.2.1) is the isomorphism in Proposition 7.1.1. The second isomorphism is the identity on  $V^*$ . The dual basis of a basis  $\alpha$  of  $V$  are the coordinates (as linear functionals) with respect to  $\alpha$ .

**Example 7.2.2.** An inner product  $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$  is a dual pairing. The induced isomorphism

$$V \cong V^*: \vec{x} \mapsto \langle \cdot, \vec{x} \rangle,$$

makes  $V$  into a *self-dual* vector space. For the dot product on  $\mathbb{R}^n$ , the self-dual isomorphism is

$$\mathbb{R}^n \cong (\mathbb{R}^n)^*: (a_1, a_2, \dots, a_n) \mapsto l(x_1, x_2, \dots, x_n) = a_1x_1 + a_2x_2 + \dots + a_nx_n.$$

More generally, a non-singular bilinear function  $b: V \times V \rightarrow \mathbb{R}$  induces two isomorphisms  $V \cong V^*$ , making  $V$  into a self-dual vector space in possibly two ways. The two ways are the same if and only if  $b(\vec{x}, \vec{y}) = b(\vec{y}, \vec{x})$ , or the bilinear function is *symmetric*. If a symmetric bilinear function further has a *self-dual basis*  $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  in the sense that

$$b(\vec{v}_i, \vec{v}_j) = \delta_{ij},$$

then the isomorphism  $V \cong \mathbb{R}^n$  induced by  $\alpha$  identifies  $b$  with the usual dot product on  $\mathbb{R}^n$ . Therefore a symmetric bilinear function with a self-dual basis is exactly an inner product. Moreover, self-dual bases with respect to the inner product are exactly orthonormal bases.

**Exercise 7.32.** Suppose  $b: V \times W \rightarrow \mathbb{R}$  is a bilinear function. Let

$$\begin{aligned} V_0 &= \{\vec{v}: b(\vec{v}, \vec{w}) = 0 \text{ for all } \vec{w} \in W\}, \\ W_0 &= \{\vec{w}: b(\vec{v}, \vec{w}) = 0 \text{ for all } \vec{v} \in V\}. \end{aligned}$$

Prove that  $b$  induces a dual pairing  $\bar{b}: \bar{V} \times \bar{W} \rightarrow \mathbb{R}$  between the quotient spaces  $\bar{V} = V/V_0$  and  $\bar{W} = W/W_0$ .

**Exercise 7.33.** Suppose  $L: V \rightarrow W$  is a linear transform of inner product spaces. Then we may combine the dual transform  $L^*: W^* \rightarrow V^*$  with the isomorphisms in Example 7.2.2 to get the *adjoint transform*  $L^*: W \rightarrow V$  with respect to the inner products (we still use the same notation as the dual transform).

1. Prove that  $L^*$  is the unique linear transform satisfying  $\langle L(\vec{x}), \vec{y} \rangle = \langle \vec{x}, L^*(\vec{y}) \rangle$ .
2. Directly verify that the adjoint satisfies the properties in Exercise 7.13.
3. For orthonormal bases  $\alpha$  and  $\beta$  of  $V$  and  $W$ , prove that  $(L^*)_{\alpha\beta} = (L_{\beta\alpha})^T$ .
4. Prove that  $\|L^*\| = \|L\|$  with respect to the norms induced by the inner products.

**Exercise 7.34.** Suppose  $b_i$  is a dual pairing between  $V_i$  and  $W_i$ ,  $i = 1, 2$ . Prove that  $b(\vec{x}_1 + \vec{x}_2, \vec{y}_1 + \vec{y}_2) = b_1(\vec{x}_1, \vec{y}_1) + b_2(\vec{x}_2, \vec{y}_2)$  is a dual pairing between the direct sums  $V_1 \oplus V_2$  and  $W_1 \oplus W_2$ .

**Exercise 7.35.** Formulate the version of Proposition 7.2.1 for  $\lim_{t \rightarrow 0} \vec{v}_t$ .

## Skew-symmetry and Cross Product

A *bilinear form* is a bilinear function on  $V \times V$ , or both variables are in the same vector space  $V$ . The bilinear form is *symmetric* if  $b(\vec{x}, \vec{y}) = b(\vec{y}, \vec{x})$ , and is *skew-symmetric* if  $b(\vec{x}, \vec{y}) = -b(\vec{y}, \vec{x})$ . Any bilinear form is the unique sum of a symmetric form and a skew-symmetric form

$$b(\vec{x}, \vec{y}) = s(\vec{x}, \vec{y}) + a(\vec{x}, \vec{y}),$$

where

$$s(\vec{x}, \vec{y}) = \frac{1}{2}(b(\vec{x}, \vec{y}) + b(\vec{y}, \vec{x})), \quad a(\vec{x}, \vec{y}) = \frac{1}{2}(b(\vec{x}, \vec{y}) - b(\vec{y}, \vec{x})).$$

**Exercise 7.36.** Prove that a bilinear form is skew-symmetric if and only if  $b(\vec{x}, \vec{x}) = 0$  for any  $\vec{x}$ .

A bilinear form on  $\mathbb{R}^n$  is

$$b(\vec{x}, \vec{y}) = \sum b_{ij} x_i y_j,$$

where  $B_{\epsilon\epsilon} = (b_{ij})$  is the matrix of the bilinear form with respect to the standard basis  $\epsilon$ . The form is symmetric if and only if  $b_{ji} = b_{ij}$ . The form is skew-symmetric if and only if  $b_{ji} = -b_{ij}$ .

A skew-symmetric bilinear form on  $\mathbb{R}^2$  is

$$b(\vec{x}, \vec{y}) = b_{12}x_1y_2 + b_{21}x_2y_1 = b_{12}x_1y_2 - b_{12}x_2y_1 = b_{12} \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}.$$

A skew-symmetric bilinear form on  $\mathbb{R}^3$  is

$$\begin{aligned} b(\vec{x}, \vec{y}) &= b_{12}x_1y_2 + b_{21}x_2y_1 + b_{13}x_1y_3 + b_{31}x_3y_1 + b_{23}x_2y_3 + b_{32}x_3y_2 \\ &= b_{12}(x_1y_2 - x_2y_1) + b_{31}(x_3y_1 - x_1y_3) + b_{23}(x_2y_3 - x_3y_2) \\ &= b_{23} \det \begin{pmatrix} x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} + b_{31} \det \begin{pmatrix} x_3 & y_3 \\ x_1 & y_1 \end{pmatrix} + b_{12} \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \\ &= \vec{b} \cdot (\vec{x} \times \vec{y}). \end{aligned}$$

In general, a skew-symmetric bilinear form on  $\mathbb{R}^n$  is

$$b(\vec{x}, \vec{y}) = \sum_{i \neq j} b_{ij}x_iy_j = \sum_{i < j} b_{ij}(x_iy_j - x_jy_i) = \sum_{i < j} b_{ij} \det \begin{pmatrix} x_i & y_i \\ x_j & y_j \end{pmatrix}. \quad (7.2.3)$$

Motivated by the formula on  $\mathbb{R}^3$ , we wish to interpret this as a dot product between vectors

$$\vec{b} = (b_{ij})_{i < j}, \quad \vec{x} \times \vec{y} = \left( \det \begin{pmatrix} x_i & y_i \\ x_j & y_j \end{pmatrix} \right)_{i < j}.$$

However, the two vectors lie in a new Euclidean space  $\mathbb{R}^{\frac{n(n-1)}{2}}$ . Here  $\frac{n(n-1)}{2}$  is the number of choices of  $(i, j)$  satisfying  $1 \leq i < j \leq n$ , and is equal to  $n$  if and only if  $n = 3$ . Therefore the  $n$ -dimensional cross product

$$\vec{x} \times \vec{y}: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^{\frac{n(n-1)}{2}}$$

should be an external operation, in the sense that the result of the operation lies out of the original vector space. Moreover, the cross products  $\vec{e}_i \times \vec{e}_j$ ,  $1 \leq i < j \leq n$ , of the standard basis vectors of  $\mathbb{R}^n$  form a standard (orthonormal) basis of the space  $\mathbb{R}^{\frac{n(n-1)}{2}}$ , making  $\mathbb{R}^{\frac{n(n-1)}{2}}$  into an inner product space. Then the skew-symmetric bilinear form  $b(\vec{x}, \vec{y})$  is the inner product between the vectors

$$\vec{b} = \sum_{i < j} b_{ij} \vec{e}_i \times \vec{e}_j, \quad \vec{x} \times \vec{y} = \sum_{i < j} \det \begin{pmatrix} x_i & y_i \\ x_j & y_j \end{pmatrix} \vec{e}_i \times \vec{e}_j.$$

The general cross product  $\vec{x} \times \vec{y}$  is still bilinear and satisfies  $\vec{x} \times \vec{y} = -\vec{y} \times \vec{x}$ . This implies  $\vec{x} \times \vec{x} = \vec{0}$  and

$$\begin{aligned} &(x_1\vec{b}_1 + x_2\vec{b}_2) \times (y_1\vec{b}_1 + y_2\vec{b}_2) \\ &= x_1y_1\vec{b}_1 \times \vec{b}_1 + x_1y_2\vec{b}_1 \times \vec{b}_2 + x_2y_1\vec{b}_2 \times \vec{b}_1 + x_2y_2\vec{b}_2 \times \vec{b}_2 \\ &= x_1y_2\vec{b}_1 \times \vec{b}_2 - x_2y_1\vec{b}_1 \times \vec{b}_2 = \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \vec{b}_1 \times \vec{b}_2. \end{aligned} \quad (7.2.4)$$

In  $\mathbb{R}^3$ , the cross product  $\vec{x} \times \vec{y}$  is a vector orthogonal to  $\vec{x}$  and  $\vec{y}$ , with the direction determined by the right hand rule from  $\vec{x}$  to  $\vec{y}$ . Moreover, the Euclidean norm of  $\vec{x} \times \vec{y}$  is given by the area of the parallelogram spanned by the two vectors

$$\|\vec{x} \times \vec{y}\|_2 = \|\vec{x}\|_2 \|\vec{y}\|_2 |\sin \theta|,$$

where  $\theta$  is the angle between the two vectors. In  $\mathbb{R}^n$ , however, we cannot compare the directions of  $\vec{x} \times \vec{y}$  and  $\vec{x}, \vec{y}$  because they lie in different vector spaces. However, the length of the cross product is still the area of the parallelogram spanned by the two vectors

$$\begin{aligned} \text{Area}(\vec{x}, \vec{y}) &= \sqrt{(\vec{x} \cdot \vec{x})(\vec{y} \cdot \vec{y}) - (\vec{x} \cdot \vec{y})^2} = \sqrt{\sum_{i,j} x_i^2 y_j^2 - \sum_{i,j} x_i y_i x_j y_j} \\ &= \sqrt{\sum_{i < j} (x_i^2 y_j^2 + x_j^2 y_i^2 - 2x_i y_i x_j y_j)} = \sqrt{\sum_{i < j} (x_i y_j - x_j y_i)^2} \\ &= \|\vec{x} \times \vec{y}\|_2. \end{aligned} \tag{7.2.5}$$

In  $\mathbb{R}^2$ , the cross product  $\vec{x} \times \vec{y} \in \mathbb{R}$  is the determinant of the matrix formed by the two vectors. The Euclidean norm of the cross product is  $|\det(\vec{x} \ \vec{y})|$ , which is the area of the parallelogram spanned by  $\vec{x}$  and  $\vec{y}$ . A linear transform  $L(\vec{x}) = x_1 \vec{a}_1 + x_2 \vec{a}_2: \mathbb{R}^2 \rightarrow \mathbb{R}^n$  takes the parallelogram spanned by  $\vec{x}, \vec{y} \in \mathbb{R}^2$  to the parallelogram spanned by  $L(\vec{x}), L(\vec{y}) \in \mathbb{R}^n$ . By (7.2.4), we have

$$\text{Area}(L(\vec{x}), L(\vec{y})) = |\det(\vec{x} \ \vec{y})| \|\vec{a}_1 \times \vec{a}_2\|_2 = \|\vec{a}_1 \times \vec{a}_2\|_2 \text{Area}(\vec{x}, \vec{y}).$$

**Exercise 7.37 (Archimedes<sup>33</sup>).** Find the area of the region  $A$  bounded by  $x = 0$ ,  $y = 4$ , and  $y = x^2$  in the following way.

1. Show that the area of the triangle with vertices  $(a, a^2)$ ,  $(a + h, (a + h)^2)$  and  $(a + 2h, (a + 2h)^2)$  is  $|h|^3$ .
2. Let  $P_n$  be the polygon bounded by the line connecting  $(0, 0)$  to  $(0, 4)$ , the line connecting  $(0, 4)$  to  $(2, 4)$ , and the line segments connecting points on the parabola  $y = x^2$  with  $x = 0, \frac{1}{2^{n-1}}, \frac{2}{2^{n-1}}, \dots, \frac{2^n}{2^{n-1}}$ .
3. Prove that the area of the polygon  $P_n - P_{n-1}$  is  $\frac{1}{4^{n-1}}$ .
4. Prove that the area of the region  $A$  is  $\sum_{n=0}^{\infty} \frac{1}{4^{n-1}} = \frac{16}{3}$ .

## Symmetry and Quadratic Form

A *quadratic form* on  $\mathbb{R}^n$  is obtained by taking two vectors in a bilinear form to be the same

$$q(\vec{x}) = b(\vec{x}, \vec{x}) = \sum_{i,j} b_{ij} x_i x_j = B\vec{x} \cdot \vec{x}.$$

<sup>33</sup>Archimedes of Syracuse, born 287 BC and died 212 BC.

Since skew-symmetric bilinear forms induce the zero quadratic form (see Exercise 7.36), we may assume that  $B$  is a symmetric matrix (i.e.,  $b$  is symmetric). Then quadratic forms are in one-to-one correspondence with symmetric matrices

$$\begin{aligned} q(x_1, x_2, \dots, x_n) &= \sum_{1 \leq i \leq n} b_{ii}x_i^2 + 2 \sum_{1 \leq i < j \leq n} b_{ij}x_i x_j \\ &= b_{11}x_1^2 + b_{22}x_2^2 + \cdots + b_{nn}x_n^2 + 2b_{12}x_1x_2 + 2b_{13}x_1x_3 + \cdots + 2b_{(n-1)n}x_{n-1}x_n. \end{aligned}$$

**Exercise 7.38.** Prove that a symmetric bilinear form can be recovered from the corresponding quadratic form by

$$b(\vec{x}, \vec{y}) = \frac{1}{4}(q(\vec{x} + \vec{y}) - q(\vec{x} - \vec{y})) = \frac{1}{2}(q(\vec{x} + \vec{y}) - q(\vec{x}) - q(\vec{y})).$$

**Exercise 7.39.** Prove that a quadratic form is homogeneous of second order

$$q(c\vec{x}) = c^2q(\vec{x}),$$

and satisfies the *parallelogram law*

$$q(\vec{x} + \vec{y}) + q(\vec{x} - \vec{y}) = 2q(\vec{x}) + 2q(\vec{y}).$$

**Exercise 7.40.** Suppose a function  $q$  satisfies the parallelogram law in Exercise 7.39. Define a function  $b(\vec{x}, \vec{y})$  by the formula in Exercise 7.38.

1. Prove that  $q(\vec{0}) = 0$ ,  $q(-\vec{x}) = q(\vec{x})$ .
2. Prove that  $b$  is symmetric.
3. Prove that  $b$  satisfies  $b(\vec{x} + \vec{y}, \vec{z}) + b(\vec{x} - \vec{y}, \vec{z}) = 2b(\vec{x}, \vec{z})$ .
4. Suppose  $f(\vec{x})$  is a function satisfying  $f(\vec{0}) = 0$  and  $f(\vec{x} + \vec{y}) + f(\vec{x} - \vec{y}) = 2f(\vec{x})$ . Prove that  $f$  is additive:  $f(\vec{x} + \vec{y}) = f(\vec{x}) + f(\vec{y})$ .
5. Prove that if  $q$  is continuous, then  $b$  is a bilinear form.

A quadratic form is *positive definite* if  $q(\vec{x}) > 0$  for any  $\vec{x} \neq 0$ . It is *negative definite* if  $q(\vec{x}) < 0$  for any  $\vec{x} \neq 0$ . If equalities are allowed, then we get *positive semi-definite* or *negative semi-definite* forms. The final possibility is *indefinite* quadratic forms, for which the value can be positive as well as negative.

A quadratic form consists of the *square terms*  $b_{ii}x_i^2$  and the *cross terms*  $b_{ij}x_i x_j$ ,  $i \neq j$ . A quadratic form without cross terms is  $q(\vec{x}) = b_{11}x_1^2 + b_{22}x_2^2 + \cdots + b_{nn}x_n^2$ , which is positive definite if and only if all  $b_{ii} > 0$ , negative definite if and only if all  $b_{ii} < 0$ , and indefinite if and only if some  $b_{ii} > 0$  and some  $b_{jj} < 0$ . For a general quadratic form, we can first eliminate the cross terms by the technique of completing the square and then determine the nature of the quadratic form.

**Example 7.2.3.** For  $q(x, y, z) = x^2 + 13y^2 + 14z^2 + 6xy + 2xz + 18yz$ , we gather together

all the terms involving  $x$  and complete the square

$$\begin{aligned} q &= x^2 + 6xy + 2xz + 13y^2 + 14z^2 + 18yz \\ &= [x^2 + 2x(3y + z) + (3y + z)^2] + 13y^2 + 14z^2 + 18yz - (3y + z)^2 \\ &= (x + 3y + z)^2 + 4y^2 + 13z^2 + 12yz. \end{aligned}$$

The remaining terms involve only  $y$  and  $z$ . Gathering all the terms involving  $y$  and completing the square, we get

$$4y^2 + 13z^2 + 12yz = (2y + 3z)^2 + 4z^2.$$

Thus  $q = (x + 3y + z)^2 + (2y + 3z)^2 + (2z)^2 = u^2 + v^2 + w^2$  and is positive definite.

**Example 7.2.4.** The cross terms in the quadratic form  $q = 4x_1^2 + 19x_2^2 - 4x_4^2 - 4x_1x_2 + 4x_1x_3 - 8x_1x_4 + 10x_2x_3 + 16x_2x_4 + 12x_3x_4$  can be eliminated as follows.

$$\begin{aligned} q &= 4[x_1^2 - x_1x_2 + x_1x_3 - 2x_1x_4] + 19x_2^2 - 4x_4^2 + 10x_2x_3 + 16x_2x_4 + 12x_3x_4 \\ &= 4 \left[ x_1^2 + 2x_1 \left( -\frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4 \right) + \left( -\frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4 \right)^2 \right] \\ &\quad + 19x_2^2 - 4x_4^2 + 10x_2x_3 + 16x_2x_4 + 12x_3x_4 - 4 \left( -\frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4 \right)^2 \\ &= 4 \left( x_1 - \frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4 \right)^2 + 18 \left[ x_2^2 + \frac{2}{3}x_2x_3 + \frac{2}{3}x_2x_4 \right] - x_3^2 - 8x_4^2 + 16x_3x_4 \\ &= (2x_1 - x_2 + x_3 - 2x_4)^2 + 18 \left[ x_2^2 + 2x_2 \left( \frac{1}{3}x_3 + \frac{1}{3}x_4 \right) + \left( \frac{1}{3}x_3 + \frac{1}{3}x_4 \right)^2 \right] \\ &\quad - x_3^2 - 8x_4^2 + 16x_3x_4 - 18 \left( \frac{1}{3}x_3 + \frac{1}{3}x_4 \right)^2 \\ &= (2x_1 - x_2 + x_3 - 2x_4)^2 + 18 \left( x_2 + \frac{1}{3}x_3 + \frac{1}{3}x_4 \right)^2 - 3(x_3^2 - 4x_3x_4) - 10x_4^2 \\ &= (2x_1 - x_2 + x_3 - 2x_4)^2 + 2(3x_2 + x_3 + x_4)^2 \\ &\quad - 3[x_3^2 + 2x_3(-2x_4) + (-2x_4)^2] - 10x_4^2 + 3(-2x_4)^2 \\ &= (2x_1 - x_2 + x_3 - 2x_4)^2 + 2(3x_2 + x_3 + x_4)^2 - 3(x_3 - 2x_4)^2 + 2x_4^2 \\ &= y_1^2 + 2y_2^2 - 3y_3^2 + 2y_4^2. \end{aligned}$$

The quadratic form is indefinite.

**Example 7.2.5.** The quadratic form  $q = 4xy + y^2$  has no  $x^2$  term. We may complete the square by using the  $y^2$  term and get  $q = (y + 2x)^2 - 4x^2 = u^2 - 4v^2$ , which is indefinite.

The quadratic form  $q = xy + yz$  has no square term. We may eliminate the cross terms by introducing  $x = x_1 + y_1$ ,  $y = x_1 - y_1$ , so that  $q = x_1^2 - y_1^2 + x_1z - y_1z$ . Then we complete the square and get  $q = \left( x_1 - \frac{1}{2}z \right)^2 - \left( y_1 + \frac{1}{2}z \right)^2 = \frac{1}{4}(x + y - z)^2 - \frac{1}{4}(x - y + z)^2$ . The quadratic form is also indefinite.

**Exercise 7.41.** Eliminate the cross terms and determine the nature of quadratic forms.



1.  $x^2 + 4xy - 5y^2$ .
2.  $2x^2 + 4xy$ .
3.  $4x_1^2 + 4x_1x_2 + 5x_2^2$ .
4.  $x^2 + 2y^2 + z^2 + 2xy - 2xz$ .
5.  $-2u^2 - v^2 - 6w^2 - 4uw + 2vw$ .
6.  $x_1^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_1x_4 + 2x_3x_4$ .

**Exercise 7.42.** Eliminate the cross terms in the quadratic form  $x^2 + 2y^2 + z^2 + 2xy - 2xz$  by first completing a square for terms involving  $z$ , then completing for terms involving  $y$ .

**Exercise 7.43.** Prove that if a quadratic form is positive definite, then the coefficients of all the square terms must be positive.

**Exercise 7.44.** Prove that if a quadratic form  $q(\vec{x})$  is positive definite, then there is  $\lambda > 0$ , such that  $q(\vec{x}) \geq \lambda \|\vec{x}\|^2$  for any  $\vec{x}$ .

Consider a quadratic form  $q(\vec{x}) = B\vec{x} \cdot \vec{x}$ , where  $B$  is a symmetric matrix. The *principal minors* of  $B$  are the determinants of the square submatrices formed by the entries in the first  $k$  rows and first  $k$  columns of  $B$

$$d_1 = b_{11}, d_2 = \det \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}, d_3 = \det \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}, \dots, d_n = \det B.$$

If  $d_1 \neq 0$ , then eliminating all the cross terms involving  $x_1$  gives

$$\begin{aligned} q(\vec{x}) &= b_{11} \left( x_1^2 + 2x_1 \frac{1}{b_{11}} (b_{12}x_2 + \dots + b_{1n}x_n) + \frac{1}{b_{11}^2} (b_{12}x_2 + \dots + b_{1n}x_n)^2 \right) \\ &\quad + b_{22}x_2^2 + \dots + b_{nn}x_n^2 + 2b_{23}x_2x_3 + 2b_{24}x_2x_4 + \dots + 2b_{(n-1)n}x_{n-1}x_n \\ &\quad - \frac{1}{b_{11}} (b_{12}x_2 + \dots + b_{1n}x_n)^2 \\ &= d_1 \left( x_1 + \frac{b_{12}}{d_1}x_2 + \dots + \frac{b_{1n}}{d_1}x_n \right)^2 + q_2(\vec{x}_2), \end{aligned}$$

where  $q_2$  is a quadratic form of the truncated vector  $\vec{x}_2 = (x_2, x_3, \dots, x_n)$ . The coefficient matrix  $B_2$  for  $q_2$  is obtained as follows. For each  $2 \leq i \leq n$ , adding  $-\frac{b_{1i}}{b_{11}}$  multiple of the first column of  $B$  to the  $i$ -th column makes the  $i$ -th term in

the first row to become zero. Then we get a matrix  $\begin{pmatrix} d_1 & \vec{0} \\ * & B_2 \end{pmatrix}$ . Since the column operation does not change the determinant of the matrix (and all the principal minors), the principal minors  $d_1^{(2)}, d_2^{(2)}, \dots, d_{n-1}^{(2)}$  of  $B_2$  are related to the principal minors  $d_1^{(1)} = d_1, d_2^{(1)} = d_2, \dots, d_n^{(1)} = d_n$  of  $B_1$  by  $d_k^{(1)} = d_1 d_k^{(2)}$ .

The discussion sets up an inductive argument. Assume  $d_1, d_2, \dots, d_k$  are all nonzero. Then we may complete the squares in  $k$  steps and obtain

$$q(\vec{x}) = d_1^{(1)}(x_1 + c_{12}x_2 + \dots + c_{1n}x_n)^2 + d_1^{(2)}(x_2 + c_{23}x_3 + \dots + c_{2n}x_n)^2 \\ + \dots + d_1^{(k)}(x_k + c_{k(k+1)}x_{k+1} + \dots + c_{kn}x_n)^2 + q_{k+1}(\vec{x}_{k+1}),$$

with

$$d_1^{(i)} = \frac{d_2^{(i-1)}}{d_1^{(i-1)}} = \frac{d_3^{(i-2)}}{d_2^{(i-2)}} = \dots = \frac{d_i^{(1)}}{d_{i-1}^{(1)}} = \frac{d_i}{d_{i-1}},$$

and the coefficient of  $x_{k+1}^2$  in  $q_{k+1}$  is  $d_1^{(k+1)} = \frac{d_{k+1}}{d_k}$ .

Thus we have shown that if  $d_1, d_2, \dots, d_n$  are all nonzero, then there is an “upper triangular” change of variables

$$\begin{aligned} y_1 &= x_1 + c_{12}x_2 + c_{13}x_3 + \dots + c_{1n}x_n, \\ y_2 &= x_2 + c_{23}x_3 + \dots + c_{2n}x_n, \\ &\vdots \\ y_n &= x_n, \end{aligned}$$

such that  $q = d_1y_1^2 + \frac{d_2}{d_1}y_2^2 + \dots + \frac{d_n}{d_{n-1}}y_n^2$ . Consequently, we get *Sylvester's criterion*.

1. If  $d_1 > 0, d_2 > 0, \dots, d_n > 0$ , then  $q(\vec{x})$  is positive definite.
2. If  $-d_1 > 0, d_2 > 0, \dots, (-1)^n d_n > 0$ , then  $q(\vec{x})$  is negative definite.
3. If  $d_1 > 0, d_2 > 0, \dots, d_k > 0, d_{k+1} < 0$ , or  $-d_1 > 0, d_2 > 0, \dots, (-1)^k d_k > 0, (-1)^{k+1} d_{k+1} < 0$ , then  $q(\vec{x})$  is indefinite.

### 7.3 Multilinear Map

A map  $F(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k): V_1 \times V_2 \times \dots \times V_k \rightarrow W$  is *multilinear* if it is linear in each of its  $k$  vector variables. For example, if  $B_1(\vec{x}, \vec{y})$  and  $B_2(\vec{u}, \vec{v})$  are bilinear, then  $B_1(\vec{x}, B_2(\vec{u}, \vec{v}))$  is a trilinear map in  $\vec{x}, \vec{u}, \vec{v}$ . The addition, scalar multiplication and composition of multilinear maps of matching types are still multilinear. If  $W = \mathbb{R}$ , then we call  $F$  a *multilinear function*. If  $W = \mathbb{R}$  and all  $V_i = V$ , then we call  $F$  a *multilinear form*.

Give bases  $\alpha_i = \{\vec{v}_{i1}, \vec{v}_{i2}, \dots, \vec{v}_{in_i}\}$  of  $V_i$ , a multilinear map is

$$F = \sum_{i_1, i_2, \dots, i_k} x_{1i_1} x_{2i_2} \dots x_{ki_k} \vec{a}_{i_1 i_2 \dots i_k}, \quad \vec{a}_{i_1 i_2 \dots i_k} = F(\vec{v}_{1i_1}, \vec{v}_{2i_2}, \dots, \vec{v}_{ki_k}).$$

Conversely, any such expression is a multilinear map. This means that a multilinear map can be defined by specifying its values at basis vectors. This also means that two multilinear maps are equal if and only if they have the same values at basis vectors.

Using the standard bases of Euclidean spaces, a multilinear map on Euclidean spaces satisfies

$$\|F(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k)\| \leq \left( \sum_{i_1, i_2, \dots, i_k} \|\vec{a}_{i_1 i_2 \dots i_k}\| \right) \|\vec{x}_1\|_\infty \|\vec{x}_2\|_\infty \cdots \|\vec{x}_k\|_\infty.$$

This implies that, for normed vector spaces  $V_i$  and  $W$ , we have

$$\|F(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k)\| \leq \lambda \|\vec{x}_1\| \|\vec{x}_2\| \cdots \|\vec{x}_k\|$$

for some constant  $\lambda$ . The smallest such  $\lambda$  as the *norm of the multilinear map*

$$\begin{aligned} \|F\| &= \inf\{\lambda: \|F(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k)\| \leq \lambda \|\vec{x}_1\| \|\vec{x}_2\| \cdots \|\vec{x}_k\| \text{ for all } \vec{x}_1, \vec{x}_2, \dots, \vec{x}_k\} \\ &= \sup\{\|F(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k)\|: \|\vec{x}_1\| = \|\vec{x}_2\| = \cdots = \|\vec{x}_k\| = 1\}. \end{aligned}$$

**Exercise 7.45.** Prove that any multilinear map from finite dimensional vector spaces is continuous.

**Exercise 7.46.** Prove that the norm of multilinear map satisfies the three conditions for norm.

**Exercise 7.47.** Study the norm of compositions such as the trilinear map  $B_1(\vec{x}, B_2(\vec{u}, \vec{v}))$ .

## High Order Form and Polynomial

A *k-form* is obtained by taking all vectors in a multilinear form  $f$  of  $k$  vectors to be the same

$$\phi(\vec{x}) = f(\vec{x}, \vec{x}, \dots, \vec{x}).$$

Similar to quadratic forms (which are 2-forms), without loss of generality, we may assume that  $f$  is *symmetric*

$$f(\vec{x}_1, \dots, \vec{x}_i, \dots, \vec{x}_j, \dots, \vec{x}_k) = f(\vec{x}_1, \dots, \vec{x}_j, \dots, \vec{x}_i, \dots, \vec{x}_k).$$

Then we have a one-to-one correspondence between  $k$ -forms and symmetric multilinear forms of  $k$  vectors.

Given a basis  $\alpha$  of  $V$ , the  $k$ -form is

$$\phi(\vec{x}) = \sum_{i_1, i_2, \dots, i_k} a_{i_1 i_2 \dots i_k} x_{i_1} x_{i_2} \cdots x_{i_k},$$

where we may further assume that the coefficient  $a_{i_1 i_2 \dots i_k} = f(\vec{v}_{i_1}, \vec{v}_{i_2}, \dots, \vec{v}_{i_k})$  is symmetric. Then the coefficient depends only on the number  $k_i$  of indices  $i_* = i$ , and we can write

$$a_{i_1 i_2 \dots i_k} x_{i_1} x_{i_2} \cdots x_{i_k} = a^{k_1 k_2 \dots k_n} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}.$$

For example, in the collection  $\{2, 4, 4\}$ , 1, 2, 3, 4 appears respectively  $k_1 = 0, k_2 = 1, k_3 = 0, k_4 = 2$  times, so that

$$a_{244}x_2x_4x_4 = a_{424}x_4x_2x_4 = a_{442}x_4x_4x_2 = a^{0102}x_1^0x_2^1x_3^0x_4^2.$$

For a quadratic form, this rephrasing leads to  $a_{ii}x_ix_i = a_{ii}x_i^2$  and  $a_{ij}x_ix_j + a_{ji}x_jx_i = 2a_{ij}x_ix_j$ . For a  $k$ -form, we get

$$\phi(\vec{x}) = \sum_{\substack{k_1+k_2+\dots+k_n=k \\ k_i \geq 0}} \frac{k!}{k_1!k_2!\dots k_n!} a^{k_1k_2\dots k_n} x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}.$$

Here  $\frac{k!}{k_1!k_2!\dots k_n!}$  is the number of ways that  $k_1$  copies of 1,  $k_2$  copies of 2,  $\dots$ ,  $k_n$  copies of  $n$ , can be arranged into ordered sequences  $i_1, i_2, \dots, i_k$ . For the case of  $k = 2, k_1 = k_2 = 1$ , this number is the scalar 2 for the cross terms in the quadratic form.

Since  $k$ -forms satisfy  $\phi(c\vec{x}) = c^k\phi(\vec{x})$ , they are homogeneous functions of order  $k$ , and are the multivariable analogue of the monomial  $x^k$  for single variable polynomials. Therefore a *multivariable polynomial* of order  $k$  is a linear combination of  $j$ -forms with  $j \leq k$

$$\begin{aligned} p(\vec{x}) &= \phi_0(\vec{x}) + \phi_1(\vec{x}) + \phi_2(\vec{x}) + \dots + \phi_k(\vec{x}) \\ &= \sum_{\substack{k_1+k_2+\dots+k_n \leq k \\ k_i \geq 0}} b^{k_1k_2\dots k_n} x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}. \end{aligned}$$

More generally, a *polynomial map*  $P: V \rightarrow W$  is

$$P(\vec{x}) = \Phi_0(\vec{x}) + \Phi_1(\vec{x}) + \Phi_2(\vec{x}) + \dots + \Phi_k(\vec{x}),$$

where

$$\Phi_l(\vec{x}) = F_l(\vec{x}, \vec{x}, \dots, \vec{x}),$$

and  $F_l: V \times V \times \dots \times V \rightarrow W$  is a multilinear map of  $l$  vectors. Without loss of generality, we may always assume that  $F_l$  is *symmetric*

$$F_l(\vec{x}_1, \dots, \vec{x}_i, \dots, \vec{x}_j, \dots, \vec{x}_l) = F_l(\vec{x}_1, \dots, \vec{x}_j, \dots, \vec{x}_i, \dots, \vec{x}_l),$$

and get a one-to-one correspondence between  $\Phi_l$  and symmetric  $F_l$ .

## Alternating Multilinear Map

A multilinear map  $F$  on  $V \times V \times \dots \times V$  is *alternating* if switching any two variables changes the sign

$$F(\vec{x}_1, \dots, \vec{x}_i, \dots, \vec{x}_j, \dots, \vec{x}_k) = -F(\vec{x}_1, \dots, \vec{x}_j, \dots, \vec{x}_i, \dots, \vec{x}_k).$$

The property is equivalent to (see Exercise 7.37)

$$F(\vec{x}_1, \dots, \vec{y}, \dots, \vec{y}, \dots, \vec{x}_k) = 0. \quad (7.3.1)$$

Given a basis  $\alpha$  of  $V$ , the alternating multilinear map is

$$F(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k) = \sum_{i_1, i_2, \dots, i_k} x_{1i_1} x_{2i_2} \cdots x_{ki_k} \vec{a}_{i_1 i_2 \cdots i_k},$$

where the coefficient  $\vec{a}_{i_1 i_2 \cdots i_k} = F(\vec{v}_{i_1}, \vec{v}_{i_2}, \dots, \vec{v}_{i_k})$  changes sign when two indices are exchanged. In particular, if  $k > n$ , then at least two indices in  $i_1, i_2, \dots, i_k$  must be the same, and the coefficient vector is zero by (7.3.1). Therefore the alternating multilinear map vanishes when  $k > n = \dim V$ .

For  $k \leq n = \dim V$ , we have

$$F(\vec{v}_{i_1}, \vec{v}_{i_2}, \dots, \vec{v}_{i_k}) = \pm F(\vec{v}_{j_1}, \vec{v}_{j_2}, \dots, \vec{v}_{j_k}),$$

where  $j_1 < j_2 < \cdots < j_k$  is the rearrangement of  $i_1, i_2, \dots, i_k$  in increasing order, and the sign  $\pm$  is the parity of the number of pair exchanges needed to recover  $(j_1, j_2, \dots, j_k)$  from  $(i_1, i_2, \dots, i_k)$ . This gives

$$\begin{aligned} F(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k) &= \sum_{i_1, i_2, \dots, i_k} x_{1i_1} x_{2i_2} \cdots x_{ki_k} F(\vec{v}_{i_1}, \vec{v}_{i_2}, \dots, \vec{v}_{i_k}) \\ &= \sum_{j_1 < j_2 < \cdots < j_k} \left( \sum_{i_* \text{ rearrange } j_*} \pm x_{1i_1} x_{2i_2} \cdots x_{ki_k} \right) F(\vec{v}_{j_1}, \vec{v}_{j_2}, \dots, \vec{v}_{j_k}) \\ &= \sum_{j_1 < j_2 < \cdots < j_k} \det \begin{pmatrix} x_{1j_1} & x_{2j_1} & \cdots & x_{kj_1} \\ x_{1j_2} & x_{2j_2} & \cdots & x_{kj_2} \\ \vdots & \vdots & & \vdots \\ x_{1j_k} & x_{2j_k} & \cdots & x_{kj_k} \end{pmatrix} F(\vec{v}_{j_1}, \vec{v}_{j_2}, \dots, \vec{v}_{j_k}). \end{aligned} \quad (7.3.2)$$

For  $k = n = \dim V$ , this shows that any multilinear alternating map of  $n$  vectors in  $n$ -dimensional space is a constant multiple of the determinant

$$F(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n) = \det \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{pmatrix} \vec{a}, \quad \vec{a} = F(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n). \quad (7.3.3)$$

For  $k = 2$  and  $V = \mathbb{R}^n$ , the formula (7.3.2) is the same as the cross product calculation in (7.2.3). Therefore (7.3.2) leads to a generalized cross product  $\vec{x}_1 \times \vec{x}_2 \times \cdots \times \vec{x}_k$  of  $k$  vectors in  $\mathbb{R}^n$ , with the value lying in a vector space with

$$\vec{e}_{i_1} \times \vec{e}_{i_2} \times \cdots \times \vec{e}_{i_k}, \quad 1 \leq i_1 < i_2 < \cdots < i_k \leq n$$

and a standard orthonormal basis. The dimension of the vector space is the number  $\frac{n!}{k!(n-k)!}$  of ways of choosing  $k$  unordered distinct numbers between 1 and  $n$ .

**Exercise 7.48.** Prove  $\det AB = \det A \det B$  by showing that both sides are multilinear alternating functions of the column vectors of  $B$ .

## Exterior Algebra

The standard mathematical language for the generalized cross product of many vectors is the *wedge product*, in which  $\times$  is replaced with  $\wedge$ . Let  $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  be a basis of  $V$ . For any subset  $I \subset [n] = \{1, 2, \dots, n\}$ , we arrange the indices in  $I$  in increasing order

$$I = \{i_1, i_2, \dots, i_k\}, \quad 1 \leq i_1 < i_2 < \dots < i_k \leq n,$$

and introduce the symbol

$$\vec{v}_{\wedge I} = \vec{v}_{i_1} \wedge \vec{v}_{i_2} \wedge \dots \wedge \vec{v}_{i_k}.$$

Define the  $k$ -th exterior power  $\Lambda^k V = \bigoplus_{|I|=k} \mathbb{R} \vec{v}_{\wedge I}$  of  $V$  to be the vector space with the collection of symbols  $\{\vec{v}_{\wedge I} : |I| = k\}$  as a basis. We have

$$\Lambda^0 V = \mathbb{R}, \quad \Lambda^1 V = V, \quad \Lambda^n V = \mathbb{R} \vec{v}_{\wedge [n]}, \quad \Lambda^k V = 0 \text{ for } k > n.$$

We also denote the basis of the top exterior power  $\Lambda^n V$  by  $\wedge \alpha = \vec{v}_{\wedge [n]}$ .

If the (ordered) indices  $(i_1, i_2, \dots, i_n)$  are not increasing, then we may change to the increasing indices  $(j_1, j_2, \dots, j_n)$  by a number of pair exchanges. We define

$$\vec{v}_{i_1} \wedge \vec{v}_{i_2} \wedge \dots \wedge \vec{v}_{i_k} = \pm \vec{v}_{j_1} \wedge \vec{v}_{j_2} \wedge \dots \wedge \vec{v}_{j_k}, \quad (7.3.4)$$

where the sign is positive if the number of pair exchanges is even and is negative if the number of pair exchanges is odd.

The *exterior algebra* of  $V$  is

$$\Lambda V = \Lambda^0 V \oplus \Lambda^1 V \oplus \Lambda^2 V \oplus \dots \oplus \Lambda^n V,$$

and has basis

$$\alpha_{\wedge} = \{\vec{v}_{\wedge I} : I \subset [n]\}.$$

It is an algebra because, in addition to the obvious vector space structure, we may define the *wedge product*

$$\wedge : \Lambda^k V \times \Lambda^l V \rightarrow \Lambda^{k+l} V, \quad \Lambda V \times \Lambda V \rightarrow \Lambda V,$$

to be the bilinear map that extends the obvious product of the basis vectors

$$\begin{aligned} \vec{v}_{\wedge I} \wedge \vec{v}_{\wedge J} &= (\vec{v}_{i_1} \wedge \dots \wedge \vec{v}_{i_k}) \wedge (\vec{v}_{j_1} \wedge \dots \wedge \vec{v}_{j_l}) \\ &= \begin{cases} \vec{v}_{i_1} \wedge \dots \wedge \vec{v}_{i_k} \wedge \vec{v}_{j_1} \wedge \dots \wedge \vec{v}_{j_l}, & \text{if } I \cap J = \emptyset, \\ \vec{0}, & \text{if } I \cap J \neq \emptyset. \end{cases} \end{aligned} \quad (7.3.5)$$

Here the definition (7.3.4) may be further used when  $(i_1, \dots, i_k, j_1, \dots, j_l)$  is not in the increasing order.

The bilinear property in the definition of the wedge product means that the product is *distributive*

$$\begin{aligned} (c_1 \vec{\lambda}_1 + c_2 \vec{\lambda}_2) \wedge \vec{\mu} &= c_1 \vec{\lambda}_1 \wedge \vec{\mu} + c_2 \vec{\lambda}_2 \wedge \vec{\mu}, \\ \vec{\lambda} \wedge (c_1 \vec{\mu}_1 + c_2 \vec{\mu}_2) &= c_1 \vec{\lambda} \wedge \vec{\mu}_1 + c_2 \vec{\lambda} \wedge \vec{\mu}_2. \end{aligned}$$

The wedge product is also *associative*

$$(\vec{\lambda} \wedge \vec{\mu}) \wedge \vec{v} = \vec{\lambda} \wedge (\vec{\mu} \wedge \vec{v}),$$

and *graded commutative*

$$\vec{\lambda} \wedge \vec{\mu} = (-1)^{kl} \vec{\mu} \wedge \vec{\lambda} \quad \text{for } \vec{\lambda} \in \Lambda^k V, \vec{\mu} \in \Lambda^l V.$$

To prove the graded commutative property, we note that both sides are bilinear for vectors in  $\Lambda^k V$  and  $\Lambda^l V$ . Therefore the equality holds if and only if it holds on the basis vectors. This reduces the proof to the equality  $\vec{v}_{\wedge I} \wedge \vec{v}_{\wedge J} = (-1)^{kl} \vec{v}_{\wedge J} \wedge \vec{v}_{\wedge I}$ , which follows from the definition (7.3.4). Similarly, the associative property can be proved by showing that the values of both sides (which are triple linear maps) are equal on basis vectors. The exterior algebra can be generated from the unit 1 and the vectors in  $V$  by the wedge product.

**Exercise 7.49.** What are the dimensions of  $\Lambda^k V$  and  $\Lambda V$ .

**Exercise 7.50.** Prove the associativity of the wedge product.

**Exercise 7.51.** Explain that the multiple wedge product map

$$\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k : V \times V \times \cdots \times V \rightarrow \Lambda^k V$$

is multilinear and alternating. Then deduce the formula of the product map in terms of the coordinates with respect to a basis

$$\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k = \sum_{i_1 < i_2 < \cdots < i_k} \det \begin{pmatrix} x_{1i_1} & x_{2i_1} & \cdots & x_{ki_1} \\ x_{1i_2} & x_{2i_2} & \cdots & x_{ki_2} \\ \vdots & \vdots & & \vdots \\ x_{1i_k} & x_{2i_k} & \cdots & x_{ki_k} \end{pmatrix} \vec{v}_{i_1} \wedge \vec{v}_{i_2} \wedge \cdots \wedge \vec{v}_{i_k}.$$

**Exercise 7.52.** Apply Exercise 7.51 to  $\mathbb{R}^n$  with the standard basis.

1. Prove that the wedge product of  $n$  vectors in  $\mathbb{R}^n$  is essentially the determinant

$$\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_n = \det(\vec{x}_1 \ \vec{x}_2 \ \cdots \ \vec{x}_n) \vec{e}_{\wedge[n]}.$$

2. Apply Exercise 7.51 to  $\vec{x}_2 \wedge \cdots \wedge \vec{x}_n$  and explain that the equality in the first part is the cofactor expansion of determinant.
3. Apply Exercise 7.51 to  $\vec{x}_1 \wedge \cdots \wedge \vec{x}_k$ ,  $\vec{x}_{k+1} \wedge \cdots \wedge \vec{x}_n$ , and get a generalization of the cofactor expansion.
4. By dividing  $\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_n$  into three or more parts, get further generalization of the cofactor expansion.

## Linear Transform of Exterior Algebra

Our construction of the exterior algebra uses a choice of basis. We will show that the exterior algebra is actually independent of the choice after studying the linear transforms of exterior algebras.

Given a linear transform  $L: V \rightarrow W$ , we define a linear transform

$$\Lambda^k L: \Lambda^k V \rightarrow \Lambda^k W$$

by linearly extending the values on the basis vectors

$$\Lambda^k L(\vec{v}_{i_1} \wedge \vec{v}_{i_2} \wedge \cdots \wedge \vec{v}_{i_k}) = L(\vec{v}_{i_1}) \wedge L(\vec{v}_{i_2}) \wedge \cdots \wedge L(\vec{v}_{i_k}), \quad i_1 < i_2 < \cdots < i_k.$$

Note that a basis  $\alpha$  of  $V$  is explicitly used here, and another basis  $\beta$  of  $W$  is implicitly used for the wedge product on the right. We may emphasize this point by writing

$$\Lambda_{\beta\alpha}^k L: \Lambda_{\alpha}^k V \rightarrow \Lambda_{\beta}^k W.$$

We may combine  $\Lambda^k L$  for all  $0 \leq k \leq n$  to form a linear transform

$$\Lambda L: \Lambda V \rightarrow \Lambda W.$$

This is in fact an algebra homomorphism because of the following properties

$$\Lambda L(1) = 1, \quad \Lambda L(a\vec{\lambda} + b\vec{\mu}) = a\Lambda L(\vec{\lambda}) + b\Lambda L(\vec{\mu}), \quad \Lambda L(\vec{\lambda} \wedge \vec{\mu}) = \Lambda L(\vec{\lambda}) \wedge \Lambda L(\vec{\mu}).$$

The first two equalities follow from the definition of  $\Lambda L$ . The third equality can be proved similar to the proof of graded commutativity. Since both  $\Lambda L(\vec{\lambda} \wedge \vec{\mu})$  and  $\Lambda L(\vec{\lambda}) \wedge \Lambda L(\vec{\mu})$  are bilinear in  $\vec{\lambda}, \vec{\mu} \in \Lambda V$ , the equality is reduced to the special case that  $\vec{\lambda}$  and  $\vec{\mu}$  are standard basis vectors

$$\Lambda L(\vec{v}_{\wedge I} \wedge \vec{v}_{\wedge J}) = \Lambda L(\vec{v}_{\wedge I}) \wedge \Lambda L(\vec{v}_{\wedge J}).$$

This can be easily verified by definition.

We also have

$$\Lambda(K \circ L) = \Lambda K \circ \Lambda L.$$

Since both sides are homomorphisms of algebras, we only need to verify the equality on generators of the exterior algebra. Since generators are simply vectors  $\vec{v} \in V$ , the only thing we need to verify is  $\Lambda(K \circ L)(\vec{v}) = (\Lambda K(\Lambda L)(\vec{v}))$ . This is simply  $(K \circ L)(\vec{v}) = K(L(\vec{v}))$ , the definition of composition.

**Exercise 7.53.** Prove that if  $L$  is injective, then  $\Lambda L$  is injective. What about subjectivity?

**Exercise 7.54.** Prove that  $\Lambda(cL) = c^k \Lambda L$  on  $\Lambda^k \mathbb{R}^n$ . Can you say anything about  $\Lambda(K + L)$ ?

Now we explain that the exterior algebra is independent of the choice of basis. Suppose  $\alpha$  and  $\alpha'$  are two bases of  $V$ . Then the identity isomorphism  $I: V \rightarrow V$  induces an isomorphism of exterior algebras

$$\Lambda_{\alpha'\alpha} I: \Lambda_{\alpha} V \rightarrow \Lambda_{\alpha'} V.$$

This identifies the two constructions  $\Lambda_{\alpha} V$  and  $\Lambda_{\alpha'} V$  of the exterior algebra. To show that the identification makes the exterior algebra truly independent of the choice



of bases, we need to further explain that identification is “natural” with respect to the exterior algebra homomorphisms induced by linear transforms. Specifically, we consider a linear transform  $L: V \rightarrow W$ , two bases  $\alpha, \alpha'$  of  $V$ , and two bases  $\beta, \beta'$  of  $W$ . We need to show that the induced algebra homomorphisms  $\Lambda_{\beta\alpha}L$  and  $\Lambda_{\beta'\alpha'}L$  correspond under the identifications of the exterior algebras. This means the equality  $\Lambda_{\beta'\alpha'}L \circ \Lambda_{\alpha'\alpha}I = \Lambda_{\beta'\beta}I \circ \Lambda_{\beta\alpha}L$ . By  $\Lambda(K \circ L) = \Lambda K \circ \Lambda L$ , both sides are equal to  $\Lambda_{\beta'\alpha}L$ , and the equality is verified.

$$\begin{array}{ccc} \Lambda_{\alpha}V & \xrightarrow{\Lambda_{\beta\alpha}L} & \Lambda_{\beta}W \\ \Lambda_{\alpha'\alpha}I \downarrow & & \downarrow \Lambda_{\beta'\beta}I \\ \Lambda_{\alpha'}V & \xrightarrow{\Lambda_{\beta'\alpha'}L} & \Lambda_{\beta'}W \end{array}$$

**Example 7.3.1.** If  $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$  are linearly independent vectors in  $V$ , then we can expand the vectors into a basis  $\alpha = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k, \vec{x}_{k+1}, \dots, \vec{x}_n\}$  of  $V$ . Since  $\vec{x}_1 \wedge \vec{x}_2 \wedge \dots \wedge \vec{x}_k$  is a vector in a basis of  $\Lambda^k V$ , we get  $\vec{x}_1 \wedge \vec{x}_2 \wedge \dots \wedge \vec{x}_k \neq \vec{0}$ . On the other hand, suppose  $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$  are linearly dependent. Then without loss of generality, we may assume

$$\vec{x}_1 = c_2 \vec{x}_2 + \dots + c_k \vec{x}_k.$$

This implies

$$\vec{x}_1 \wedge \vec{x}_2 \wedge \dots \wedge \vec{x}_k = c_2 \vec{x}_2 \wedge \vec{x}_2 \wedge \dots \wedge \vec{x}_k + \dots + c_k \vec{x}_k \wedge \vec{x}_2 \wedge \dots \wedge \vec{x}_k = \vec{0}.$$

We conclude that  $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$  are linearly independent if and only if  $\vec{x}_1 \wedge \vec{x}_2 \wedge \dots \wedge \vec{x}_k \neq \vec{0}$ .

**Exercise 7.55.** Prove that if  $\vec{\lambda} \in \Lambda^k V$  is nonzero, then there is  $\vec{\mu} \in \Lambda^{n-k} V$ ,  $n = \dim V$ , such that  $\vec{\lambda} \wedge \vec{\mu} \neq \vec{0}$ .

Since the top exterior power  $\Lambda^n V$  is 1-dimensional, the linear transform  $\Lambda^n L$  induced from  $L: V \rightarrow V$  is simply multiplying a number. The number is the *determinant* of the linear transform

$$\Lambda^n L(\vec{\lambda}) = (\det L) \vec{\lambda}, \quad \vec{\lambda} \in \Lambda^n V.$$

The equality can also be written as

$$L(\vec{x}_1) \wedge L(\vec{x}_2) \wedge \dots \wedge L(\vec{x}_n) = (\det L) \vec{x}_1 \wedge \vec{x}_2 \wedge \dots \wedge \vec{x}_n.$$

Moreover, the equality  $\Lambda(K \circ L) = \Lambda K \circ \Lambda L$  implies

$$\det(K \circ L) = \det K \det L.$$

**Exercise 7.56.** Suppose

$$\begin{aligned} \vec{y}_1 &= a_{11} \vec{x}_1 + a_{12} \vec{x}_2 + \dots + a_{1k} \vec{x}_k, \\ \vec{y}_2 &= a_{21} \vec{x}_1 + a_{22} \vec{x}_2 + \dots + a_{2k} \vec{x}_k, \\ &\vdots \\ \vec{y}_k &= a_{k1} \vec{x}_1 + a_{k2} \vec{x}_2 + \dots + a_{kk} \vec{x}_k, \end{aligned}$$

and denote the coefficient matrix  $A = (a_{ij})$ . Prove that

$$\vec{y}_1 \wedge \vec{y}_2 \wedge \cdots \wedge \vec{y}_k = (\det A) \vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k.$$

**Exercise 7.57.** Prove that the determinant of a linear transform  $L: V \rightarrow V$  is equal to  $\det L_{\alpha\alpha}$  of the matrix of the linear transform.

**Exercise 7.58.** Prove that  $\wedge\alpha = (\det I_{\beta\alpha})(\wedge\beta)$  and  $\wedge\beta^* = (\det I_{\beta\alpha})(\wedge\alpha^*)$ .

**Exercise 7.59.** Use the determinant defined by the exterior algebra to explain that

$$L(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = (x_1, \dots, x_i + cx_j, \dots, x_j, \dots, x_n): \mathbb{R}^n \rightarrow \mathbb{R}^n$$

has determinant 1. What about the other two elementary operations?

**Exercise 7.60.** Let  $K: V \rightarrow W$  and  $L: W \rightarrow V$  be linear transforms. Prove that if  $\dim V = \dim W$ , then  $\det(L \circ K) = \det(K \circ L)$ . What happens if  $\dim V \neq \dim W$ ?

## Dual of Exterior Algebra

Let  $b: V \times W \rightarrow \mathbb{R}$  be a dual pairing. Let  $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$  and  $\beta = \{\vec{w}_1, \dots, \vec{w}_n\}$  be dual bases of  $V$  and  $W$  with respect to  $b$ . Then define a bilinear function  $\Lambda b: \Lambda V \times \Lambda W \rightarrow \mathbb{R}$  by

$$\Lambda b(\vec{v}_{\wedge I}, \vec{w}_{\wedge J}) = \delta_{I,J}. \quad (7.3.6)$$

This means that the induced bases  $\alpha_{\wedge}$  and  $\beta_{\wedge}$  form dual bases of  $\Lambda V$  and  $\Lambda W$  with respect to  $\Lambda b$ . In particular,  $\Lambda b$  is a dual pairing.

The dual pairing decomposes into dual pairings  $\Lambda^k b: \Lambda^k V \times \Lambda^k W \rightarrow \mathbb{R}$  that satisfies

$$\Lambda^k b(\vec{x}_1 \wedge \cdots \wedge \vec{x}_k, \vec{y}_1 \wedge \cdots \wedge \vec{y}_k) = \det(b(\vec{x}_i, \vec{y}_j))_{1 \leq i, j \leq k}. \quad (7.3.7)$$

The reason is that both sides are multilinear functions of  $2k$  variables in  $V$  and  $W$ , and the equality holds when  $\vec{x}_i$  and  $\vec{y}_j$  are basis vectors in  $\alpha$  and  $\beta$ . The equality (7.3.7) implies that, if  $\alpha'$  and  $\beta'$  is another pair of dual bases of  $V$  and  $W$  with respect to  $b$ , then  $\alpha'_{\wedge}$  and  $\beta'_{\wedge}$  also form dual bases of  $\Lambda V$  and  $\Lambda W$  with respect to  $\Lambda b$ . In particular,  $\Lambda b$  is actually independent of the choice of dual bases, although its construction makes explicit use of  $\alpha$  and  $\beta$ .

For the special case that  $b$  is the evaluation pairing  $\langle \vec{x}, l \rangle = l(\vec{x}): V \times V^* \rightarrow \mathbb{R}$  in Example 7.2.1, we get a natural dual pairing between the exterior algebras  $\Lambda V$  and  $\Lambda V^*$ . This gives a natural isomorphism

$$\Lambda^k V^* = (\Lambda^k V)^*,$$

that sends  $l_1 \wedge \cdots \wedge l_k \in \Lambda^k V^*$  to the following linear functional on  $\Lambda^k V$

$$\begin{aligned} (l_1 \wedge \cdots \wedge l_k)(\vec{x}_1 \wedge \cdots \wedge \vec{x}_k) &= \langle \vec{x}_1 \wedge \cdots \wedge \vec{x}_k, l_1 \wedge \cdots \wedge l_k \rangle \\ &= \det(\langle \vec{x}_i, l_j \rangle) = \det(l_j(\vec{x}_i)). \end{aligned}$$

For any linear functional  $l \in (\Lambda^k V)^*$ ,  $l(\vec{x}_1 \wedge \cdots \wedge \vec{x}_k)$  is a multilinear alternating function. Conversely, for any multilinear alternating function  $f$  on  $V$ , let  $l$  be the linear function on  $\Lambda^k V$  uniquely determined by its values on basis vectors

$$l(\vec{v}_{i_1} \wedge \cdots \wedge \vec{v}_{i_k}) = f(\vec{v}_{i_1}, \dots, \vec{v}_{i_k}).$$

Then we have

$$f(\vec{x}_1, \dots, \vec{x}_k) = l(\vec{x}_1 \wedge \cdots \wedge \vec{x}_k),$$

because both sides are multilinear functions with the same value on basis vectors. Therefore the dual space  $(\Lambda V)^*$  is exactly the vector space of all multilinear alternating functions on  $V$ .

As an example, the determinant of  $n$  vectors in  $\mathbb{R}^n$  is a multilinear alternating function and can be regarded as the linear functional on the top exterior power  $\Lambda^n \mathbb{R}^n$  satisfying  $\det(\vec{e}_{\wedge[n]}) = 1$ . This shows that

$$\det = \wedge \epsilon^* = \vec{e}_{\wedge[n]}^* = \vec{e}_1^* \wedge \cdots \wedge \vec{e}_n^* \in \Lambda^n (\mathbb{R}^n)^* \quad (7.3.8)$$

is the dual basis  $\wedge \epsilon^*$  of the standard basis  $\wedge \epsilon = \vec{e}_{\wedge[n]}$  of  $\Lambda^n \mathbb{R}^n$ . For a general  $n$ -dimensional vector space  $V$ , therefore, we may regard any nonzero linear functional in  $\Lambda^n V^*$  as a determinant of  $n$  vectors in  $V$ . In fact, any basis  $\alpha$  of  $V$  gives a determinant  $\det_\alpha = \wedge \alpha^* \in \Lambda^n V^*$ .

**Exercise 7.61.** For a linear transform  $L: V \rightarrow W$ , prove that the dual of the transform  $\Lambda L: \Lambda V \rightarrow \Lambda W$  can be identified with  $\Lambda L^*: \Lambda W^* \rightarrow \Lambda V^*$ .

**Exercise 7.62.** For two bases  $\alpha, \beta$  of  $V$ , what is the relation between  $\det_\alpha, \det_\beta \in \Lambda^n V^*$ ?

**Exercise 7.63.** Suppose  $\phi \in \Lambda^k V^*$  and  $\psi \in \Lambda^l V^*$ . Prove that

$$\phi \wedge \psi(\vec{x}_1 \wedge \cdots \wedge \vec{x}_{k+l}) = \sum \pm \phi(\vec{x}_{i_1} \wedge \cdots \wedge \vec{x}_{i_k}) \psi(\vec{x}_{j_1} \wedge \cdots \wedge \vec{x}_{j_l}),$$

where the sum runs over all the decompositions

$$\{1, \dots, k+l\} = \{i_1, \dots, i_k\} \cup \{j_1, \dots, j_l\}, \quad i_1 < \cdots < i_k, \quad j_1 < \cdots < j_l,$$

and the sign is the parity of the number of pair exchanges needed to rearrange  $(i_1, \dots, i_k, j_1, \dots, j_l)$  into ascending order.

**Exercise 7.64.** Let  $\vec{y} \in V$ . For  $\phi \in \Lambda^k V^*$ , define  $i_{\vec{y}}\phi \in \Lambda^{k-1} V^*$  by

$$i_{\vec{y}}\phi(\vec{x}_1 \wedge \cdots \wedge \vec{x}_{k-1}) = \phi(\vec{y} \wedge \vec{x}_1 \wedge \cdots \wedge \vec{x}_{k-1}).$$

1. Use the relation between multilinear alternating functions and elements of  $\Lambda V^*$  to explain that the formula indeed defines an element  $i_{\vec{y}}\phi \in \Lambda^{k-1} V^*$ .
2. Show that  $\phi = \psi$  if and only if  $i_{\vec{y}}\phi = i_{\vec{y}}\psi$  for all  $\vec{y}$ .
3. Use Exercise 7.63 to show that

$$i_{\vec{y}}(\phi \wedge \psi) = i_{\vec{y}}\phi \wedge \psi + (-1)^k \phi \wedge i_{\vec{y}}\psi \text{ for } \phi \in \Lambda^k V^*, \psi \in \Lambda^l V^*.$$

4. Extend the discussion to  $i_\eta: \Lambda^k V^* \rightarrow \Lambda^{k-p} V^*$  for  $\eta \in \Lambda^p V$ .

**Exercise 7.65.** Prove that a bilinear map (not necessarily a dual pairing)  $b: V \times W \rightarrow \mathbb{R}$  induces a bilinear map  $\Lambda^k b: \Lambda^k V \times \Lambda^k W \rightarrow \mathbb{R}$ , such that (7.3.7) holds. Moreover, extend Exercise 7.63.

**Exercise 7.66.** Suppose  $b_i$  is a dual pairing between  $V_i$  and  $W_i$ ,  $i = 1, 2$ . Exercise 7.34 gives a dual pairing between  $V_1 \oplus V_2$  and  $W_1 \oplus W_2$ . Prove that for  $\vec{\lambda}_i \in \Lambda^{k_i} V_i$ ,  $\vec{\mu}_i \in \Lambda^{k_i} W_i$ , we have  $\Lambda^{k_1+k_2} b(\vec{\lambda}_1 \wedge \vec{\lambda}_2, \vec{\mu}_1 \wedge \vec{\mu}_2) = \Lambda^{k_1} b_1(\vec{\lambda}_1, \vec{\mu}_1) \Lambda^{k_2} b_2(\vec{\lambda}_2, \vec{\mu}_2)$ .

**Exercise 7.67.** Consider multilinear alternating functions  $f$  and  $g$  on  $V$  and  $W$  as elements of  $\Lambda^k V^* \subset \Lambda^k(V \oplus W)^*$  and  $\Lambda^l W^* \subset \Lambda^l(V \oplus W)^*$ . Then we have the multilinear alternating function  $f \wedge g$  on  $V \oplus W$ . Prove that if each  $\vec{x}_i$  is in either  $V$  or  $W$ , then  $f \wedge g(\vec{x}_1, \dots, \vec{x}_{k+l}) = 0$  unless  $k$  of  $\vec{x}_i$  are in  $V$  and  $l$  of  $\vec{x}_i$  are in  $W$ . Moreover, prove that if  $\vec{x}_i \in V$  for  $1 \leq i \leq k$  and  $\vec{x}_i \in W$  for  $k+1 \leq i \leq k+l$ , then  $f \wedge g(\vec{x}_1, \dots, \vec{x}_{k+l}) = f(\vec{x}_1, \dots, \vec{x}_k)g(\vec{x}_{k+1}, \dots, \vec{x}_{k+l})$ .

Next we apply the dual pairing of exterior algebra to an inner product  $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$  in Example 7.2.2. A basis of  $V$  is self dual if and only if it is orthonormal. Therefore an orthonormal basis  $\alpha$  of  $V$  induces a self dual basis  $\alpha_\wedge$  with respect to the induced dual pairing  $\langle \cdot, \cdot \rangle: \Lambda V \times \Lambda V \rightarrow \mathbb{R}$ . This implies that the induced dual pairing is an inner product on the exterior algebra. The equality (7.3.7) gives an explicit formula for this inner product

$$\langle \vec{x}_1 \wedge \dots \wedge \vec{x}_k, \vec{y}_1 \wedge \dots \wedge \vec{y}_k \rangle = \det(\langle \vec{x}_i, \vec{y}_j \rangle)_{1 \leq i, j \leq k}. \quad (7.3.9)$$

For a subspace  $W \subset V$ , an orthonormal basis  $\alpha = \{\vec{v}_1, \dots, \vec{v}_k\}$  of  $W$  can be extended to an orthonormal basis  $\beta = \{\vec{v}_1, \dots, \vec{v}_n\}$  of  $V$ . Since the natural map  $\Lambda W \rightarrow \Lambda V$  takes an orthonormal basis  $\alpha_\wedge$  of  $\Lambda W$  injectively into an orthonormal basis  $\beta_\wedge$  of  $\Lambda V$ , we find that  $\Lambda W$  is a subspace of  $\Lambda V$ , and the inclusion  $\Lambda W \subset \Lambda V$  preserves the inner product.

**Exercise 7.68.** Prove that if  $U$  is an orthogonal transform (i.e., preserving inner product) of an inner product space  $V$ , then  $\Lambda U$  is an orthogonal transform of  $\Lambda V$ .

**Exercise 7.69.** Suppose  $W$  is a subspace of an inner product space  $V$ , and  $W^\perp$  is the orthogonal complement of  $W$  in  $V$ . Prove that  $\Lambda W$  and  $\Lambda W^\perp$  are orthogonal in  $\Lambda V$ . Moreover, for  $\vec{\lambda} \in \Lambda W$  and  $\vec{\eta} \in \Lambda W^\perp$ , prove that  $\langle \vec{\lambda} \wedge \vec{\mu}, \vec{\xi} \wedge \vec{\eta} \rangle = \langle \vec{\lambda}, \vec{\xi} \rangle \langle \vec{\mu}, \vec{\eta} \rangle$ .

## 7.4 Orientation

### Orientation of Vector Space

Two ordered bases  $\alpha$  and  $\beta$  of a vector space  $V$  are related by a matrix  $I_{\beta\alpha}$ . They are *compatibly oriented* if  $\det I_{\beta\alpha} > 0$ .

We emphasize that the order of vectors in the basis is critical for the orientation, because if  $\beta$  is obtained from  $\alpha$  by switching two vectors, then  $\det I_{\beta\alpha} = -1$ .

Fix one (ordered) basis  $\alpha$ . Then

$$\{\text{all (ordered) bases}\} = o_\alpha \sqcup (-o_\alpha), \quad (7.4.1)$$

where

$$o_\alpha = \{\beta: \det I_{\beta\alpha} > 0\}, \quad -o_\alpha = \{\beta: \det I_{\beta\alpha} < 0\}.$$

By  $I_{\gamma\alpha} = I_{\gamma\beta}I_{\beta\alpha}$  (see Exercise 7.7), we have  $\det I_{\gamma\alpha} = \det I_{\gamma\beta} \det I_{\beta\alpha}$ . Then we may use the equality to show that any two  $\beta, \gamma \in o_\alpha$  are compatibly oriented, any two in  $-o_\alpha$  are also compatibly oriented. Moreover, the equality implies that any  $\beta \in o_\alpha$  and  $\gamma \in -o_\alpha$  are never compatibly oriented. Therefore the decomposition (7.4.1) is the equivalence classes defined by compatible orientation. This interpretation shows that the decomposition is independent of the choice of  $\alpha$ , except that the labeling of the two subsets by the signs  $\pm$  depends on  $\alpha$ . See Exercises 7.70 and 7.71.

**Definition 7.4.1.** An *orientation*  $o$  of a vector space  $V$  is a choice of one of two collections of compatibly oriented ordered bases of  $V$ . The other collection  $-o$  is then the *opposite* of the orientation  $o$ .

A basis  $\alpha$  gives the orientation  $o_\alpha$ . The standard basis  $\epsilon = \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$  gives the (*standard*) *positive orientation* on  $\mathbb{R}^n$ . This means the rightward direction of  $\mathbb{R}$ , the counterclockwise rotation in  $\mathbb{R}^2$  and the right hand rule on  $\mathbb{R}^3$ . The basis  $\{-\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$  gives the (*standard*) *negative orientation* of  $\mathbb{R}^n$ . This means the leftward direction on  $\mathbb{R}$ , the clockwise rotation in  $\mathbb{R}^2$  and the left hand rule on  $\mathbb{R}^3$ .

In an oriented vector space, the bases in the orientation collection  $o$  are *positively oriented*, and those in the opposite orientation collection  $-o$  are *negatively oriented*.

A basis  $\alpha$  of an  $n$ -dimensional vector space  $V$  gives a nonzero element  $\wedge\alpha \in \Lambda^n V - \vec{0}$ . If  $\beta$  is another basis, then by Exercise 7.57, we have  $\wedge\alpha = (\det I_{\beta\alpha})(\wedge\beta)$ . Therefore  $\alpha$  and  $\beta$  are compatibly oriented if and only if  $\wedge\alpha$  and  $\wedge\beta$  lie in the same connected component of  $\Lambda^n V - \vec{0} \cong \mathbb{R} - 0$ . In other words, an orientation of  $V$  is equivalent to a choice of one of two components of  $\Lambda^n V - \vec{0}$ . We also use  $o$  to denote the chosen component of  $\Lambda^n V - \vec{0}$ . For example, we have

$$o_\alpha = \{t(\wedge\alpha): t > 0\} \subset \Lambda^n V - \vec{0}.$$

**Exercise 7.70.** Prove that  $\det I_{\beta\alpha} > 0$  implies  $o_\alpha = o_\beta$  and  $-o_\alpha = -o_\beta$ , and prove that  $\det I_{\beta\alpha} < 0$  implies  $o_\alpha = -o_\beta$  and  $-o_\alpha = o_\beta$ . This implies that the decomposition (7.4.1) is independent of the choice of  $\alpha$ .

**Exercise 7.71.** Prove that the compatible orientation is an equivalence relation, and (7.4.1) is the corresponding decomposition into equivalence classes.

**Exercise 7.72.** Determine which operations on bases preserve the orientation, and which reverse the orientation.

1. Exchange two basis vectors.
2. Multiply a nonzero number to a basis vector.

3. Add a scalar multiple of one basis vector to another basis vector.

**Exercise 7.73.** Determine which is positively oriented, and which is negatively oriented.

- |  |   |
|--|---|
| 1. $\{\vec{e}_2, \vec{e}_3, \dots, \vec{e}_n, \vec{e}_1\}$ . | 4. $\{(1, 2), (3, 4)\}$ .                   |
| 2. $\{\vec{e}_n, \vec{e}_{n-1}, \dots, \vec{e}_1\}$ .        | 5. $\{(1, 2, 3), (4, 5, 6), (7, 8, 10)\}$ . |
| 3. $\{-\vec{e}_1, -\vec{e}_2, \dots, -\vec{e}_n\}$ .         | 6. $\{(1, 0, 1), (0, 1, 1), (1, 1, 0)\}$ .  |

If a basis gives an orientation  $o$  of  $V$ , then its dual basis gives the dual orientation  $o^*$  of  $V^*$ . By Exercise 7.16, it is easy to show that the definition is independent of the choice of basis. In fact, if a basis  $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  of  $V$  is positively oriented, then a basis  $\beta = \{l_1, l_2, \dots, l_n\}$  of  $V^*$  is positively oriented if and only if  $\det(l_j(\vec{v}_i)) > 0$  (see Exercise 7.76 and the second part of Exercise 7.78).

Let  $L: V \rightarrow W$  be an isomorphism of vector spaces. If  $V$  and  $W$  are oriented with  $o_V$  and  $o_W$ , then  $L$  *preserves the orientation* if it maps  $o_V$  to  $o_W$ . We also say that  $L$  *reverses the orientation* if it maps  $o_V$  to  $-o_W$ . On the other hand, if  $o_V$  (or  $o_W$ ) is given, then  $L$  *translates* the orientation to  $o_W$  (or  $o_V$ ) by requiring that  $L$  preserves the orientation.

If  $L: V \rightarrow V$  is an isomorphism of  $V$  to itself, then  $L$  *preserves the orientation* if  $\det L > 0$ . This means that  $L$  maps any orientation  $o$  of  $V$  to the same orientation  $o$ . The concept does not require  $V$  to be already oriented.

**Exercise 7.74.** Prove that an isomorphism  $L: V \rightarrow W$  preserves orientations  $o_V$  and  $o_W$  if and only if  $\Lambda^n L$  maps the component  $o_V \subset \Lambda^n V - \vec{0}$  to the component  $o_W \subset \Lambda^n W - \vec{0}$ .

**Exercise 7.75.** Prove that if isomorphisms  $K$  and  $L$  preserve orientation, then  $K \circ L$  and  $L^{-1}$  preserve orientation. What if both reverse orientation? What if one preserves and the other reverses orientation?

**Exercise 7.76.** Prove that the dual orientation of an orientation  $o \subset \Lambda^n V - \vec{0}$  is given by

$$o^* = \{l \in \Lambda^n V^*: l(\xi) > 0 \text{ for any } \xi \in o\} = \{l \in \Lambda^n V^*: l(\xi_0) > 0 \text{ for one } \xi_0 \in o\}.$$

**Exercise 7.77.** Suppose an isomorphism  $L: V \rightarrow W$  translates  $o_V$  to  $o_W$ . Prove that  $L^*: W^* \rightarrow V^*$  translates  $o_W^*$  to  $o_V^*$ .

**Exercise 7.78.** Suppose  $b$  is a dual pairing between  $V$  and  $W$ . Suppose  $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  and  $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$  are dual bases of  $V$  and  $W$  with respect to  $b$ . Define orientations  $o_V$  and  $o_W$  of  $V$  and  $W$  to be dual orientations with respect to  $b$  if  $\alpha \in o_V$  is equivalent to  $\beta \in o_W$ .

1. Prove that the definition of dual orientation with respect to  $b$  is independent of the choice of dual bases.
2. Prove that  $o_\alpha$  and  $o_\beta$  are dual orientations if and only if  $\det(b(\vec{v}_i, \vec{w}_j)) > 0$ .
3. Prove that  $o_V$  and  $o_W$  are dual orientations with respect to  $b$  if and only if the induced isomorphism  $V \cong W^*$  translates  $o_V$  and  $o_W^*$ .

**Exercise 7.79.** Suppose  $V$  is an inner product space. By considering the inner product as a pairing of  $V$  with itself, we may apply Exercise 7.78. Prove that the dual of any orientation  $o$  with respect to the inner product is  $o$  itself.

**Exercise 7.80.** Does an orientation of a vector space  $V$  naturally induces an orientation of  $\Lambda^k V$ ?

**Exercise 7.81.** Suppose  $V = V_1 \oplus V_2$ . Suppose  $o_1$  and  $o_2$  are orientations of  $V_1$  and  $V_2$ . We may choose an orientation compatible basis of  $V_1$  followed by an orientation compatible basis of  $V_2$  to define an orientation  $o$  of  $V$ . Prove that  $o$  is well defined. How is the orientation  $o$  changed if we exchange the order of  $V_1$  and  $V_2$ ? Can orientations of  $V_1$  and  $V$  determine an orientation of  $V_2$ ?

## Volume

The exterior algebra is closely related to the volume. In fact, the determinant of a matrix can be interpreted as a combination of the volume of the parallelotope spanned by the volume vectors (the absolute value of the determinant), and the orientation of volume vectors (the sign of the determinant).

A *parallelotope*<sup>34</sup> spanned by vectors  $\alpha = \{\vec{x}_1, \dots, \vec{x}_k\}$  is

$$P_\alpha = \{c_1\vec{x}_1 + \dots + c_k\vec{x}_k : 0 \leq c_i \leq 1\}.$$

By volume, we mean translation invariant measures on vector spaces. By Theorem 11.4.4, such measures are unique up to multiplying (positive) constant (the constant is similar to the ratio between litre and gallon). The constant may be determined by the measure of any one parallelotope spanned by a basis.

**Theorem 7.4.2.** Suppose  $\mu$  is a translation invariant measure on an  $n$ -dimensional vector space  $V$ . Then there is  $l \in \Lambda^n V^*$ , such that for any  $n$  vectors  $\alpha$  in  $V$ , we have

$$\mu(P_\alpha) = |l(\wedge \alpha)| = |l(\vec{x}_1 \wedge \dots \wedge \vec{x}_n)|.$$

The theorem basically says that there is a one-to-one correspondence between translation invariant measures on  $V$  and elements of

$$|\Lambda^n V^*| - 0 = \frac{\Lambda^n V^* - 0}{l \sim -l} \cong (0, +\infty).$$

*Proof.* Let  $\beta = \{\vec{v}_1, \dots, \vec{v}_n\}$  be a basis of  $V$ , let  $\beta^* = \{\vec{v}_1^*, \dots, \vec{v}_n^*\}$  be the dual basis, and let

$$l = \mu(P_\beta)(\wedge \beta^*) = \mu(P_\beta)\vec{v}_1^* \wedge \dots \wedge \vec{v}_n^*.$$

We prove that  $f(\alpha) = \mu(P_\alpha)$  is equal to  $g(\alpha) = |l(\wedge \alpha)|$  by studying the effect of elementary column operations on  $f$  and  $g$ .

<sup>34</sup>A parallelogram is a 2-dimensional parallelotope. A parallelepiped is a 3-dimensional parallelotope.

The first operation is the exchange of two vectors

$$\alpha = \{\vec{v}_1, \dots, \vec{v}_i, \dots, \vec{v}_j, \dots, \vec{v}_n\} \mapsto \alpha' = \{\vec{v}_1, \dots, \vec{v}_j, \dots, \vec{v}_i, \dots, \vec{v}_n\}.$$

Since  $P_{\alpha'} = P_{\alpha}$ , we get  $f(\alpha') = f(\alpha)$ . Moreover, by  $\wedge \alpha' = -\wedge \alpha$ , we get  $g(\alpha') = g(\alpha)$ .

The second is multiplying a number to one vector

$$\alpha = \{\vec{v}_1, \dots, \vec{v}_i, \dots, \vec{v}_n\} \mapsto \alpha' = \{\vec{v}_1, \dots, c\vec{v}_i, \dots, \vec{v}_n\}.$$

The parallelotopes  $P_{\alpha'}$  and  $P_{\alpha}$  have the same base parallelotope spanned by  $\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_{i+1}, \dots, \vec{v}_k$ , but the  $\vec{v}_i$  direction of  $P_{\alpha'}$  is  $c$  multiple of  $P_{\alpha}$ . Therefore the volume is multiplied by  $|c|$ , and we get  $f(\alpha') = |c|f(\alpha)$ . Moreover, by  $\wedge \alpha' = c(\wedge \alpha)$ , we get  $g(\alpha') = |c|g(\alpha)$ .

The third is adding a multiple of one vector to another

$$\alpha = \{\vec{v}_1, \dots, \vec{v}_i, \dots, \vec{v}_j, \dots, \vec{v}_n\} \mapsto \alpha' = \{\vec{v}_1, \dots, \vec{v}_i + c\vec{v}_j, \dots, \vec{v}_j, \dots, \vec{v}_n\}.$$

The parallelotopes  $P_{\alpha'}$  and  $P_{\alpha}$  have the same base parallelotope spanned by  $\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_{i+1}, \dots, \vec{v}_k$ , and the  $\vec{v}_i$  direction is shifted to  $\vec{v}_i + c\vec{v}_j$ . The shift does not change the “height” of the parallelotope as measured from the base. Therefore the two parallelotopes have the same volume, and we get  $f(\alpha') = f(\alpha)$ . Moreover, by  $\wedge \alpha' = \wedge \alpha$ , we get  $g(\alpha') = g(\alpha)$ .

We conclude that the effects of the three column operations on  $f$  and  $g$  are the same. If  $\alpha$  is not a basis, then  $f(\alpha) = 0$  since  $P_{\alpha}$  is collapsed to be of dimension  $< n$ , and  $g(\alpha) = 0$  by Example 7.3.1. If  $\alpha$  is a basis, then repeatedly applying three operations reduces  $\alpha$  to the basis  $\beta$ . Therefore the equality  $f(\alpha) = g(\alpha)$  is reduced to  $f(\beta) = g(\beta)$ . This is true because

$$g(\beta) = |l(\wedge \beta)| = \mu(P_{\beta}) = f(\beta). \quad \square$$

If  $V$  is an inner product space, then the standard Lebesgue measure  $\mu_L$  on  $V$  is the translation invariant measure such that the volume of any unit cube is 1. Since a unit cube is the parallelotope spanned by any orthonormal basis  $\beta = \{\vec{v}_1, \dots, \vec{v}_n\}$ , by the proof above, we may take  $l = \vec{v}_1^* \wedge \dots \wedge \vec{v}_n^*$ , so that

$$\mu_L(P_{\{\vec{x}_1, \dots, \vec{x}_n\}}) = |\langle \vec{x}_1 \wedge \dots \wedge \vec{x}_n, \vec{v}_1^* \wedge \dots \wedge \vec{v}_n^* \rangle| = |\det(\vec{v}_i^*(\vec{x}_j))|.$$

If  $V = \mathbb{R}^n$  and  $\beta$  is the standard basis, then by (7.3.8), this is the absolute value of the determinant of the matrix with  $\vec{x}_i$  as volume vectors.

The inner product on  $V$  induces an inner product and the associated norm on  $\Lambda^n V$ , and  $l: \Lambda^n V \cong \mathbb{R}$  is an isomorphism that preserves the norm. Therefore we also get

$$\mu_L(P_{\{\vec{x}_1, \dots, \vec{x}_n\}}) = |l(\vec{x}_1 \wedge \dots \wedge \vec{x}_n)| = \|\vec{x}_1 \wedge \dots \wedge \vec{x}_n\|_2.$$

Now consider the special case that  $V$  is a subspace of  $\mathbb{R}^m$ , with the inner product inherited from the standard dot product in  $\mathbb{R}^m$ . Then the inner product preserving embedding  $V \subset \mathbb{R}^m$  induces an inner product preserving embedding  $\Lambda^n V \subset \Lambda^n \mathbb{R}^m$ . This means that the norm  $\|\vec{x}_1 \wedge \dots \wedge \vec{x}_n\|_2$  in  $\Lambda^n V$  is also the norm in  $\Lambda^n \mathbb{R}^m$ . Replacing  $m, n$  by  $n, k$ , we get the following statement about the size of any parallelotope in Euclidean space.



**Proposition 7.4.3.** *The  $k$ -dimensional volume of the parallelotope  $P_\alpha$  spanned by  $k$  vectors  $\alpha = \{\vec{x}_1, \dots, \vec{x}_k\}$  in  $\mathbb{R}^n$  is*

$$\|\vec{x}_1 \wedge \dots \wedge \vec{x}_k\|_2 = \sqrt{\det(\langle \vec{x}_i, \vec{x}_j \rangle)_{1 \leq i, j \leq k}}.$$

Now we additionally assume that  $V$  is oriented, which means a preferred choice  $o^*$  of two components of  $\Lambda^n V^* - 0$ . Then we may choose  $l \in o^*$  in Theorem 7.4.2 and get

$$l(\wedge \alpha) = l(\vec{x}_1 \wedge \dots \wedge \vec{x}_n) = \begin{cases} \mu(P_\alpha), & \text{if } \alpha \in o, \\ -\mu(P_\alpha), & \text{if } \alpha \notin o, \\ 0, & \text{if } \alpha \text{ is not a basis.} \end{cases}$$

This shows that  $l(\wedge \alpha)$  is the *signed volume* of the parallelotope  $P_\alpha$  that takes into account of the orientation of  $\alpha$ . For the special case that  $V = \mathbb{R}^n$  with standard orientation and  $l = \vec{e}_1^* \wedge \dots \wedge \vec{e}_n^*$ , the signed volume is exactly the determinant.

**Exercise 7.82.** Suppose  $V$  is an oriented inner product space, and  $\alpha$  and  $\beta$  are positively oriented orthonormal bases. Prove that  $\wedge \alpha = \wedge \beta$ . This shows that the signed volume  $l = \wedge \alpha^*$  is independent of the choice of  $\alpha$ .

## Hodge Dual

Let  $V$  be an  $n$ -dimensional vector space. Let  $e$  be a nonzero vector of the 1-dimensional vector space  $\Lambda^n V$ , inducing an isomorphism  $\Lambda^n V \cong_e \mathbb{R}$ . Then for any  $k \leq n$ , we have a bilinear function

$$b_e: \Lambda^k V \times \Lambda^{n-k} V \xrightarrow{\wedge} \Lambda^n V \cong_e \mathbb{R}, \quad \vec{\lambda} \wedge \vec{\mu} = b_e(\vec{\lambda}, \vec{\mu})e.$$

Exercise 7.55 shows that this is a dual pairing.

Suppose  $V$  is an oriented inner product space and  $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$  is a positively oriented orthonormal basis. We take  $e = \wedge \alpha = \vec{v}_1 \wedge \dots \wedge \vec{v}_n$ , which by Exercise 7.82 is independent of the choice of  $\alpha$ . In fact,  $e$  can be characterized as the unique unit length vector lying in the orientation component  $o_V \subset \Lambda^n V - \vec{0}$ . Moreover, the inner product on  $V$  induces an inner product on  $\Lambda V$ , which further gives an isomorphism  $(\Lambda^{n-k} V)^* \cong \Lambda^{n-k} V$ . Combined with the dual pairing above, we get the *Hodge dual* (note the distinction between  $\star$  and  $*$ )

$$\vec{\lambda} \mapsto \vec{\lambda}^\star: \Lambda^k V \cong (\Lambda^{n-k} V)^* \cong \Lambda^{n-k} V.$$

This means that

$$\vec{\lambda} \wedge \vec{\mu} = b_e(\vec{\lambda}, \vec{\mu})e = \langle \vec{\lambda}^\star, \vec{\mu} \rangle e.$$

For the oriented orthonormal basis  $\alpha$ , we have

$$\langle \vec{v}_{\wedge I}^\star, \vec{v}_{\wedge J} \rangle e = \vec{v}_{\wedge I} \wedge \vec{v}_{\wedge J} = \begin{cases} \vec{0}, & \text{if } J \neq [n] - I, \\ \vec{v}_{\wedge I} \wedge \vec{v}_{\wedge ([n] - I)}, & \text{if } J = [n] - I. \end{cases}$$

For fixed  $I$ ,  $\vec{v}_{\wedge I}^*$  is the unique vector such that the equality above holds for all  $J$ . This leads to  $\vec{v}_{\wedge I}^* = \pm \vec{v}_{\wedge([n]-I)}$ , where the sign  $\pm$  is determined by

$$\vec{v}_{\wedge I} \wedge \vec{v}_{\wedge([n]-I)} = \pm e = \pm \vec{v}_1 \wedge \cdots \wedge \vec{v}_n.$$

For example, we have

$$\begin{aligned} 1^* &= \vec{v}_{\wedge[n]}, \\ \vec{v}_i^* &= (-1)^{i-1} \vec{v}_{\wedge([n]-i)}, \\ \vec{v}_{i \wedge j}^* &= (-1)^{i+j-1} \vec{v}_{\wedge([n]-\{i,j\})} \quad \text{for } i < j, \\ \vec{v}_{\wedge([n]-i)}^* &= (-1)^{n-i} \vec{v}_i, \\ \vec{v}_{\wedge[n]}^* &= 1. \end{aligned}$$

The calculation shows that the Hodge dual operation sends the orthonormal basis  $\alpha_\wedge$  of  $\Lambda V$  to essentially itself (up to exchanging orders and adding signs). Therefore the Hodge dual preserves the inner product and the associated length

$$\langle \vec{\lambda}, \vec{\mu} \rangle = \langle \vec{\lambda}^*, \vec{\mu}^* \rangle, \quad \|\vec{\lambda}^*\| = \|\vec{\lambda}\|.$$

In  $\mathbb{R}^2$  (with counterclockwise orientation), the Hodge dual is the counterclockwise rotation by 90 degrees

$$(x_1, x_2)^* = x_1 \vec{e}_1^* + x_2 \vec{e}_2^* = x_1 \vec{e}_2 - x_2 \vec{e}_1 = (-x_2, x_1),$$

The usual cross product in  $\mathbb{R}^3$  is the Hodge dual of the wedge product

$$\begin{aligned} &[(x_1, x_2, x_3) \wedge (y_1, y_2, y_3)]^* \\ &= \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} (\vec{e}_1 \wedge \vec{e}_2)^* + \det \begin{pmatrix} x_1 & y_1 \\ x_3 & y_3 \end{pmatrix} (\vec{e}_1 \wedge \vec{e}_3)^* + \det \begin{pmatrix} x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} (\vec{e}_2 \wedge \vec{e}_3)^* \\ &= \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \vec{e}_3 - \det \begin{pmatrix} x_1 & y_1 \\ x_3 & y_3 \end{pmatrix} \vec{e}_2 + \det \begin{pmatrix} x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} \vec{e}_1 \\ &= (x_1, x_2, x_3) \times (y_1, y_2, y_3). \end{aligned}$$

**Exercise 7.83.** For  $\vec{\lambda} \in \Lambda^k V$ , prove that

$$\vec{\lambda}^{**} = (-1)^{k(n-k)} \vec{\lambda}, \quad \vec{\lambda} \wedge \vec{\mu}^* = \langle \vec{\lambda}, \vec{\mu} \rangle e, \quad \vec{\lambda} \wedge \vec{\mu} = (-1)^{k(n-k)} \langle \vec{\lambda}, \vec{\mu}^* \rangle e.$$

**Exercise 7.84.** Suppose an isomorphism  $L: V \rightarrow W$  preserves the orientation and the inner product. Prove that  $\Lambda L(\vec{\lambda})^* = \Lambda L(\vec{\lambda}^*)$ . What happens when  $L$  reverses the orientation?

**Exercise 7.85.** Suppose  $W$  is a subspace of an inner product space  $V$  and  $W^\perp$  is the orthogonal complement of  $W$  in  $V$ . Suppose  $W$  and  $W^\perp$  are oriented, and  $V$  is compatibly oriented by Exercise 7.81. Prove that  $e_W^* = e_{W^\perp}$ .

**Exercise 7.86.** Suppose  $\alpha = \{\vec{x}_1, \dots, \vec{x}_n\}$  is a positively oriented basis of  $\mathbb{R}^n$ , such that the first  $k$  vectors in  $\alpha$  and the last  $n-k$  vectors are orthogonal. Prove that

$$(\vec{x}_1 \wedge \cdots \wedge \vec{x}_k)^* = \frac{\det(\vec{x}_1 \cdots \vec{x}_n)}{\|\vec{x}_{k+1} \wedge \cdots \wedge \vec{x}_n\|_2^2} \vec{x}_{k+1} \wedge \cdots \wedge \vec{x}_n. \quad (7.4.2)$$

Exercise 7.87. The cross product in  $\mathbb{R}^n$  is the map

$$\vec{x}_1 \times \cdots \times \vec{x}_{n-1} = (\vec{x}_1 \wedge \cdots \wedge \vec{x}_{n-1})^*: \mathbb{R}^n \times \cdots \times \mathbb{R}^n \rightarrow \mathbb{R}^n.$$

Find the explicit formula and show that the cross product is orthogonal to every  $\vec{x}_i$ .

Exercise 7.88. Prove that  $(\vec{x}_1 \times \cdots \times \vec{x}_{n-1}) \cdot \vec{x}_n = \det(\vec{x}_1 \cdots \vec{x}_n)$  in  $\mathbb{R}^n$ , and explain that this is the cofactor expansion of determinant.

## 7.5 Additional Exercises

### Isometry between Normed Spaces

Let  $\|\vec{x}\|$  and  $\|\vec{y}\|$  be norms on  $\mathbb{R}^n$  and  $\mathbb{R}^m$ . A map  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is an *isometry* if it satisfies  $\|F(\vec{x}) - F(\vec{y})\| = \|\vec{x} - \vec{y}\|$ . A map  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is *affine* if the following equivalent conditions are satisfied.

1.  $F((1-t)\vec{x} + t\vec{y}) = (1-t)F(\vec{x}) + tF(\vec{y})$  for any  $\vec{x}, \vec{y} \in \mathbb{R}^n$  and any  $0 \leq t \leq 1$ .
2.  $F((1-t)\vec{x} + t\vec{y}) = (1-t)F(\vec{x}) + tF(\vec{y})$  for any  $\vec{x}, \vec{y} \in \mathbb{R}^n$  and any  $t \in \mathbb{R}$ .
3.  $F(\vec{x}) = \vec{a} + L(\vec{x})$  for a fixed vector  $\vec{a}$  (necessarily equal to  $F(\vec{0})$ ) and a linear transform  $L$ .

Exercise 7.89. Suppose the norm  $\|\vec{x}\|$  on  $\mathbb{R}^m$  is *strictly convex*, in the sense  $\|\vec{x} + \vec{y}\| = \|\vec{x}\| + \|\vec{y}\|$  implies  $\vec{x}$  and  $\vec{y}$  are parallel. Prove that for any isometry  $F$  and  $\vec{x}, \vec{y}, \vec{z} = (1-t)\vec{x} + t\vec{y} \in \mathbb{R}^n$ ,  $0 \leq t \leq 1$ , the vectors  $F(\vec{z}) - F(\vec{x})$  and  $F(\vec{y}) - F(\vec{z})$  must be parallel. Then prove that the isometry  $F$  is affine.

Exercise 7.90. Show that the  $L^p$ -norm is strictly convex for  $1 < p < \infty$ . Then show that an isometry between Euclidean spaces with the Euclidean norms must be of the form  $\vec{a} + L(\vec{x})$ , where  $L$  is a linear transform with its matrix  $A$  satisfying  $A^T A = I$ .

Exercise 7.91. For any  $\vec{u} \in \mathbb{R}^n$ , the map  $\phi(\vec{x}) = 2\vec{u} - \vec{x}$  is the reflection with respect to  $\vec{u}$ . A subset  $K \subset \mathbb{R}^n$  is symmetric with respect to  $\vec{u}$  if  $\vec{x} \in K$  implies  $\phi(\vec{x}) \in K$ . The subset has radius  $r(K) = \sup_{\vec{x} \in K} \|\vec{x} - \vec{u}\|$ .

1. Prove that  $\phi$  is an isometry from  $(\mathbb{R}^n, \|\cdot\|)$  to itself,  $\|\phi(\vec{x}) - \vec{x}\| = 2\|\vec{x} - \vec{u}\|$ , and  $\vec{u}$  is the only point fixed by  $\phi$  (i.e., satisfying  $\phi(\vec{x}) = \vec{x}$ ).
2. For a subset  $K$  symmetric with respect to  $\vec{u}$ , prove that the subset has diameter  $\sup_{\vec{x}, \vec{y} \in K} \|\vec{x} - \vec{y}\| = 2r(K)$ . Then prove that the subset  $K' = \{\vec{x}: K \subset B(\vec{x}, r(K))\}$  has radius  $r(K') \leq \frac{1}{2}r(K)$ .
3. For any  $\vec{a}, \vec{b} \in \mathbb{R}^n$ , denote  $\vec{u} = \frac{\vec{a} + \vec{b}}{2}$ . Prove that

$$K_0 = \{\vec{x}: \|\vec{x} - \vec{a}\| = \|\vec{x} - \vec{b}\| = \frac{1}{2}\|\vec{a} - \vec{b}\|\}$$

is symmetric with respect to  $\vec{u}$ . Then prove that the sequence  $K_n$  defined by  $K_{n+1} = K'_n$  satisfies  $\cap K_n = \{\vec{u}\}$ .

The last part gives a characterization of the middle point  $\vec{u}$  of two points  $\vec{a}$  and  $\vec{b}$  purely in terms of the norm.

**Exercise 7.92 (Mazur-Ulam Theorem).** Prove that an invertible isometry is necessarily affine. Specifically, suppose  $F: (\mathbb{R}^n, \|\vec{x}\|) \rightarrow (\mathbb{R}^n, \|\vec{x}\|)$  is an invertible isometry. By using the characterization of the middle point in Exercise 7.90, prove that the map preserves the middle point

$$F\left(\frac{\vec{a} + \vec{b}}{2}\right) = \frac{F(\vec{a}) + F(\vec{b})}{2}.$$

Then further prove that the property implies  $F$  is affine.

**Exercise 7.93.** Let  $\phi(t)$  be a real function and consider  $F(t) = (t, \phi(t)): \mathbb{R} \rightarrow \mathbb{R}^2$ . For the absolute value on  $\mathbb{R}$  and the  $L^\infty$ -norm on  $\mathbb{R}^2$ , find suitable condition on  $\phi$  to make sure  $F$  is an isometry. The exercise shows that an isometry is not necessarily affine in general.

## Chapter 8

# Multivariable Differentiation

## 8.1 Linear Approximation

The differentiation of maps between Euclidean spaces (or more generally, finite dimensional vector spaces) can be defined by directly generalizing the definition for single variable functions. Denote  $\Delta\vec{x} = \vec{x} - \vec{x}_0$ . Any linear map can be expressed as  $\vec{a} + L(\Delta\vec{x})$  for a constant vector  $\vec{a}$  and a linear transform  $L$ .

**Definition 8.1.1.** A map  $F(\vec{x})$  defined on a ball around  $\vec{x}_0$  is *differentiable* at  $\vec{x}_0$  if there is a linear map  $\vec{a} + L(\Delta\vec{x})$ , such that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|\Delta\vec{x}\| < \delta \implies \|F(\vec{x}) - \vec{a} - L(\Delta\vec{x})\| \leq \epsilon \|\Delta\vec{x}\|.$$

The linear transform  $L$  is the *derivative* of  $F$  at  $\vec{x}_0$  and is denoted  $F'(\vec{x}_0)$ .

Like the single variable case, the definition can be rephrased as

$$\vec{a} = F(\vec{x}_0), \quad \lim_{\Delta\vec{x} \rightarrow \vec{0}} \frac{\|F(\vec{x}_0 + \Delta\vec{x}) - F(\vec{x}_0) - L(\Delta\vec{x})\|}{\|\Delta\vec{x}\|} = 0.$$

Equivalently, we may write

$$F(\vec{x}) = \vec{a} + L(\Delta\vec{x}) + o(\|\Delta\vec{x}\|) = F(\vec{x}_0) + F'(\vec{x}_0)(\Delta\vec{x}) + o(\|\Delta\vec{x}\|),$$

or

$$\Delta F = F(\vec{x}) - F(\vec{x}_0) = F'(\vec{x}_0)(\Delta\vec{x}) + o(\|\Delta\vec{x}\|).$$

The differentiability at a point requires the map to be defined everywhere near the point. Therefore the differentiability is defined for maps on open subsets. In the future, the definition may be extended to maps on “differentiable subsets” (called submanifolds).

Similar to the single variable case, we denote the *differential*  $dF = L(d\vec{x})$  of a map  $F$ . Again at the moment it is only a symbolic notation.

A single variable function is a map  $f: \mathbb{R} \rightarrow \mathbb{R}$ . Its derivative is a linear transform  $f'(x_0): \mathbb{R} \rightarrow \mathbb{R}$ . However, a linear transform from  $\mathbb{R}$  to itself is always a multiplication by a number. The number corresponding to the linear transform  $f'(x_0)$  is the derivative, also denoted  $f'(x_0)$ .

**Example 8.1.1.** For the multiplication map  $\mu(x, y) = xy: \mathbb{R}^2 \rightarrow \mathbb{R}$ , we have

$$\mu(x, y) = (x_0 + \Delta x)(y_0 + \Delta y) = x_0 y_0 + y_0 \Delta x + x_0 \Delta y + \Delta x \Delta y.$$

Since  $y_0 \Delta x + x_0 \Delta y$  is linear in  $(\Delta x, \Delta y)$ , and  $|\Delta x \Delta y| \leq \|(\Delta x, \Delta y)\|_\infty^2$ , we see that the multiplication map is differentiable, the derivative linear transform is  $(u, v) \mapsto y_0 u + x_0 v$ , and the differential is  $d_{(x_0, y_0)}(xy) = y_0 dx + x_0 dy$ .

**Example 8.1.2.** For the map  $F(x, y) = (x^2 + y^2, xy): \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , we have

$$\begin{aligned} F(x, y) - F(x_0, y_0) &= (2x_0 \Delta x + \Delta x^2 + 2y_0 \Delta y + \Delta y^2, y_0 \Delta x + x_0 \Delta y + \Delta x \Delta y) \\ &= (2x_0 \Delta x + 2y_0 \Delta y, y_0 \Delta x + x_0 \Delta y) + (\Delta x^2 + \Delta y^2, \Delta x \Delta y). \end{aligned}$$

Since  $(2x_0\Delta x + 2y_0\Delta y, y_0\Delta x + x_0\Delta y)$  is a linear transform of  $(\Delta x, \Delta y)$ , and  $\|(\Delta x^2 + \Delta y^2, \Delta x\Delta y)\|_\infty \leq 2\|(\Delta x, \Delta y)\|_\infty^2$ , the map  $F$  is differentiable at  $(x_0, y_0)$ , and the derivative  $F'(x_0, y_0)$  is the linear transform  $(u, v) \mapsto (2x_0u + 2y_0v, y_0u + x_0v)$ . The differential is  $d_{(x_0, y_0)}F = (2x_0dx + 2y_0dy, y_0dx + x_0dy)$ .

**Example 8.1.3.** For the function  $f(\vec{x}) = \vec{x} \cdot \vec{x} = \|\vec{x}\|_2^2$ , we have

$$\|\vec{x}_0 + \Delta\vec{x}\|_2^2 = \|\vec{x}_0\|_2^2 + 2\vec{x}_0 \cdot \Delta\vec{x} + \|\Delta\vec{x}\|_2^2.$$

Since  $2\vec{x}_0 \cdot \Delta\vec{x}$  is a linear functional of  $\Delta\vec{x}$ , and  $\|\Delta\vec{x}\|_2^2 = o(\|\Delta\vec{x}\|_2)$ , the function  $f$  is differentiable, with the derivative  $f'(\vec{x}_0)(\vec{v}) = 2\vec{x}_0 \cdot \vec{v}$  and the differential  $d_{\vec{x}_0}f = 2\vec{x}_0 \cdot d\vec{x}$ .

**Example 8.1.4.** A curve in a vector space  $V$  is a map  $\phi: (a, b) \rightarrow V$ . The differentiability at  $t_0$  means that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|\Delta t| < \delta \implies \|\phi(t) - \vec{a} - \Delta t \vec{b}\| \leq \epsilon |\Delta t|.$$

By the same proof as Proposition 3.1.5, this is equivalent to that  $\phi$  is continuous at  $t_0$ ,  $\vec{a} = \phi(t_0)$ , and the limit

$$\phi'(t_0) = \vec{b} = \lim_{t \rightarrow t_0} \frac{\phi(t) - \phi(t_0)}{t - t_0}$$

converges. Note that in Definition 8.1.1, the notation  $\phi'(t_0)$  is used for the derivative as a *linear transform* from  $\mathbb{R} \rightarrow V$ . In the limit above, the same notation is used as the *tangent vector* of the curve. They are related by

$$\begin{aligned} [\text{tangent vector } \phi'(t_0)] &= [\text{linear transform } \phi'(t_0)](1), \\ [\text{linear transform } \phi'(t_0)](t) &= t[\text{tangent vector } \phi'(t_0)]. \end{aligned}$$

The relation takes advantage of the fact that a linear transform  $L: \mathbb{R} \rightarrow V$  is determined by the vector  $L(1) \in V$ .

**Example 8.1.5.** The space of  $n \times n$  matrices form a vector space  $M(n) \cong \mathbb{R}^{n^2}$ . For the map  $F(X) = X^2: \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$  of the square of matrices, we have

$$\Delta F = F(A + H) - F(A) = (A^2 + AH + HA + H^2) - A^2 = AH + HA + H^2.$$

The map  $H \mapsto AH + HA$  is a linear transform. Moreover, by Proposition 7.1.2, we have  $\|H^2\| \leq \|H\|^2 = o(\|H\|)$ . Therefore the map is differentiable, with the derivative  $F'(A)(H) = AH + HA$  and the differential  $d_AF = A(dX) + (dX)A$ .

**Example 8.1.6.** To find the derivative of the determinant  $\det: M(n) \rightarrow \mathbb{R}$  at the identity matrix  $I$ , we need to find the linear approximation of  $\det(I + H)$ . The  $i$ -th column of  $I$  is the standard basis vector  $\vec{e}_i$ . Let the  $i$ -th column of  $H$  be  $\vec{h}_i$ . Since the determinant is linear in each of its column vectors, we have

$$\begin{aligned} \det(\vec{e}_1 + \vec{h}_1 \quad \vec{e}_2 + \vec{h}_2 \quad \cdots \quad \vec{e}_n + \vec{h}_n) &= \det(\vec{e}_1 \quad \vec{e}_2 \quad \cdots \quad \vec{e}_n) \\ &\quad + \sum_{1 \leq i \leq n} \det(\vec{e}_1 \quad \cdots \quad \vec{h}_i \quad \cdots \quad \vec{e}_n) \\ &\quad + \sum_{1 \leq i < j \leq n} \det(\vec{e}_1 \quad \cdots \quad \vec{h}_i \quad \cdots \quad \vec{h}_j \quad \cdots \quad \vec{e}_n) \\ &\quad + \cdots \end{aligned}$$

The sum starts at  $\det(\vec{e}_1 \ \vec{e}_2 \ \cdots \ \vec{e}_n) = \det I = 1$  and gradually replaces  $\vec{e}_i$  with  $\vec{h}_i$ . The linear terms consist of  $\det(\vec{e}_1 \ \cdots \ \vec{h}_i \ \cdots \ \vec{e}_n) = h_{ii}$ , in which only one  $\vec{e}_i$  is replaced by  $\vec{h}_i$ . Therefore the derivative is the sum of diagonal entries (called the *trace*)

$$\det'(I)(H) = \text{tr}H = \sum_{1 \leq i \leq n} h_{ii}.$$

**Exercise 8.1.** Use the definition to show the differentiability of  $ax^2 + 2bxy + cy^2$  and find the derivative.

**Exercise 8.2.** Prove that if a map is differentiable at  $\vec{x}_0$ , then the map is continuous at  $\vec{x}_0$ . Then show that the Euclidean norm  $\|\vec{x}\|_2$  is continuous but not differentiable at  $\vec{0}$ .

**Exercise 8.3.** Suppose a map  $F$  is differentiable at  $\vec{x}_0$ , with  $F(\vec{x}_0) = \vec{0}$ . Suppose a function  $\lambda(\vec{x})$  is continuous at  $\vec{x}_0$ . Prove that  $\lambda(\vec{x})F(\vec{x})$  is differentiable at  $\vec{x}_0$ .

**Exercise 8.4.** Prove that a function  $f(\vec{x})$  is differentiable at  $\vec{x}_0$  if and only if  $f(\vec{x}) = f(\vec{x}_0) + J(\vec{x}) \cdot \Delta\vec{x}$ , where  $J: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is continuous at  $\vec{x}_0$ . Extend the fact to differentiable maps.

**Exercise 8.5.** A function  $f(\vec{x})$  is *homogeneous* of degree  $p$  if  $f(c\vec{x}) = c^p f(\vec{x})$  for any  $c > 0$ . Find the condition for the function to be differentiable at  $\vec{0}$ .

**Exercise 8.6.** Suppose  $A$  is an  $n \times n$  matrix. Find the derivative of the function  $A\vec{x} \cdot \vec{x}$ .

**Exercise 8.7.** Suppose  $B: U \times V \rightarrow W$  is a bilinear map. Prove that the  $B$  is differentiable, with the derivative

$$B'(\vec{x}_0, \vec{y}_0)(\vec{u}, \vec{v}) = B(\vec{u}, \vec{y}_0) + B(\vec{x}_0, \vec{v}): U \times V \rightarrow W.$$

Extend the result to multilinear maps. This extends Example 8.1.1 and will be further extended in Exercise 8.31.

**Exercise 8.8.** Suppose  $B: U \times V \rightarrow W$  is a bilinear map, and  $\phi(t)$  and  $\psi(t)$  are differentiable curves in  $U$  and  $V$ . Then  $B(\phi(t), \psi(t))$  is a curve in  $W$ . Prove the Leibniz rule (see more general version in Exercise 8.31)

$$(B(\phi(t), \psi(t)))' = B(\phi'(t), \psi(t)) + B(\phi(t), \psi'(t)).$$

Here the derivatives are the tangent vectors in Example 8.1.4. Moreover, extend the Leibniz rule to multilinear maps.

**Exercise 8.9.** Use Proposition 7.2.1 and Exercise 7.35 to prove that, if  $b$  is a dual pairing between  $V$  and  $W$ , and  $\phi(t)$  is a differentiable curve in  $V$ , then  $\phi'(t_0) = \vec{v}$  (using tangent vector in Example 8.1.4) if and only if

$$\left. \frac{d}{dt} \right|_{t=t_0} b(\phi(t), \vec{w}) = b(\vec{v}, \vec{w}) \text{ for all } \vec{w} \in W.$$



**Exercise 8.10.** Let  $X^T$  be the transpose of a matrix  $X$ . Prove that the derivative of the map  $F(X) = X^T X: M(n) \rightarrow M(n)$  is the linear transform

$$F'(A)(H) = A^T H + H^T A.$$

**Exercise 8.11.** Use the idea of Example 8.1.6 to find the derivative of the determinant at any matrix  $A$ . Moreover, show that if  $A$  is an  $n \times n$  matrix of rank  $\leq n - 2$ , then the derivative is zero.

**Exercise 8.12.** For any natural number  $k$ , find the derivative of the map of taking the  $k$ -th power of matrices.

**Exercise 8.13.** Use the equality  $(I + H)^{-1} = I - H + H^2(I + H)^{-1}$  and Exercise 7.21 to find the derivative of the inverse matrix map at the identity matrix  $I$ .

## Partial Derivative

Let  $F(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \dots, f_m(\vec{x}))$ . By taking the  $L^\infty$ -norm, the definition for the differentiability of  $F$  means that for each  $i$ , we have

$$\|\Delta \vec{x}\| < \delta \implies |f_i(\vec{x}) - a_i - l_i(\Delta \vec{x})| \leq \epsilon \|\Delta \vec{x}\|,$$

where  $l_i$  is the  $i$ -th coordinate of the linear transform  $L$ . Therefore the map is differentiable if and only if each coordinate function is differentiable. Moreover, the linear approximation  $L$  of the map is obtained by putting together the linear approximations  $l_i$  of the coordinate functions.

A function  $f(\vec{x}) = f(x_1, x_2, \dots, x_n)$  is approximated by a linear function

$$a + l(\Delta \vec{x}) = a + b_1 \Delta x_1 + b_2 \Delta x_2 + \dots + b_n \Delta x_n$$

at  $\vec{x}_0 = (x_{10}, x_{20}, \dots, x_{n0})$  if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\begin{aligned} |\Delta x_i| &= |x_i - x_{i0}| < \delta \text{ for all } i \\ \implies |f(x_1, x_2, \dots, x_n) - a - b_1 \Delta x_1 - b_2 \Delta x_2 - \dots - b_n \Delta x_n| \\ &\leq \epsilon \max\{|\Delta x_1|, |\Delta x_2|, \dots, |\Delta x_n|\}. \end{aligned}$$

If we fix  $x_2 = x_{20}, \dots, x_n = x_{n0}$ , and let only  $x_1$  change, then the above says that  $f(x_1, x_{20}, \dots, x_{n0})$  is approximated by the linear function  $a + b_1 \Delta x_1$  at  $x_1 = x_{10}$ . The coefficients are  $a = f(\vec{x}_0)$  and the *partial derivative* of  $f(\vec{x})$  in  $x_1$

$$b_1 = \lim_{\Delta x_1 \rightarrow 0} \frac{f(x_{10} + \Delta x_1, x_{20}, \dots, x_{n0}) - f(x_{10}, x_{20}, \dots, x_{n0})}{\Delta x_1}.$$

The other coefficients are the similar partial derivatives and denoted

$$b_i = \frac{\partial f}{\partial x_i} = D_{x_i} f = f_{x_i}.$$

Using the notation, the derivative  $f'(\vec{x})$  is the linear functional

$$f'(\vec{x})(\vec{v}) = \frac{\partial f}{\partial x_1} v_1 + \frac{\partial f}{\partial x_2} v_2 + \dots + \frac{\partial f}{\partial x_n} v_n = \nabla f(\vec{x}) \cdot \vec{v}: \mathbb{R}^n \rightarrow \mathbb{R},$$

where

$$\nabla f = \left( \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

is the *gradient* of the function. Moreover, the differential of the function is

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \dots + \frac{\partial f}{\partial x_n} dx_n = \nabla f \cdot d\vec{x}.$$

Any linear functional on a finite dimensional inner product space is of the form  $l(\vec{x}) = \vec{a} \cdot \vec{x}$  for a unique vector  $\vec{a}$ . The gradient  $\nabla f(\vec{x})$  is the unique vector associated to the derivative linear functional  $f'(\vec{x})$ .

Putting the linear approximations of  $f_i$  together, we get the linear approximation of  $F$ . This means that the derivative linear transform  $F'(\vec{x}_0)$  is given by the *Jacobian matrix*

$$\frac{\partial F}{\partial \vec{x}} = \frac{\partial(f_1, f_2, \dots, f_m)}{\partial(x_1, x_2, \dots, x_n)} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

A single variable function is a map  $f(x): \mathbb{R} \rightarrow \mathbb{R}$ . Its Jacobian matrix is a  $1 \times 1$  matrix, which is the same as a number. This number is the usual derivative  $f'(x_0)$ .

**Example 8.1.7.** The function  $f(x, y) = 1 + 2x + xy^2$  has the following partial derivatives at  $(0, 0)$

$$f(0, 0) = 1, \quad f_x(0, 0) = (2 + y^2)|_{x=0, y=0} = 2, \quad f_y(0, 0) = 2xy|_{x=0, y=0} = 0.$$

Therefore the candidate for the linear approximation is  $1 + 2(x - 0) + 0(y - 0) = 1 + 2x$ . However, for the differentiability, we still need to verify the linear approximation. By  $\Delta x = x$ ,  $\Delta y = y$ , we get  $|f(x, y) - 1 - 2x| = |xy^2| \leq \|(x, y)\|_\infty^3$ . This verifies that the candidate is indeed a linear approximation, so that  $f$  is differentiable at  $(0, 0)$ , with  $d_{(0,0)}f = 2dx$ .

**Example 8.1.8.** The function  $f(x, y) = \sqrt{|xy|}$  satisfies  $f(0, 0) = 0$ ,  $f_x(0, 0) = 0$ ,  $f_y(0, 0) = 0$ . Therefore the candidate for the linear approximation at  $(0, 0)$  is the zero function. However, by Example 6.2.2, the limit

$$\lim_{(x,y) \rightarrow (0,0)} \frac{|f(x, y)|}{\|(x, y)\|_2} = \lim_{x \rightarrow 0, y \rightarrow 0} \sqrt{\frac{|xy|}{x^2 + y^2}}$$

diverges. We conclude that  $f$  is not differentiable at  $(0, 0)$ , despite the existence of the partial derivatives.

**Example 8.1.9.** For the map  $F$  in Example 8.1.2, we have  $F = (f, g)$ , where  $f = x^2 + y^2$  and  $g = xy$ . By  $f_x = 2x$ ,  $f_y = 2y$ ,  $g_x = y$ ,  $g_y = x$ , we find that, if  $F$  is differentiable, then the matrix for the derivative linear transform  $F'$  is  $\begin{pmatrix} 2x & 2y \\ y & x \end{pmatrix}$ . Indeed, this is exactly the matrix for the linear transform in Example 8.1.2. However, as we saw in Example 8.1.8, the computation of the partial derivatives does not yet imply the differentiability and is therefore not a substitute for the argument in the earlier example.

**Exercise 8.14.** Discuss the existence of partial derivatives and the differentiability at  $(0, 0)$ . Assume all the parameters are positive.

$$1. f(x, y) = \begin{cases} (x^2 + y^2)^p \sin \frac{1}{x^2 + y^2}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

$$2. f(x, y) = \begin{cases} |x|^p |y|^q \sin \frac{1}{x^2 + y^2}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

$$3. f(x, y) = \begin{cases} \frac{|x|^p |y|^q}{(|x|^m + |y|^n)^k}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

$$4. f(x, y) = \begin{cases} \frac{(|x|^p + |y|^q)^r}{(|x|^m + |y|^n)^k}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

**Exercise 8.15.** Use the equality  $f'(\vec{x})(\vec{v}) = \nabla f \cdot \vec{v}$  and Example 8.1.3 to compute the gradient of the function  $\vec{x} \cdot \vec{x}$ . The key point here is that you are not supposed to use the coordinates or partial derivatives.

**Exercise 8.16.** What is the gradient of the determinant of  $2 \times 2$  matrix?

**Exercise 8.17.** Express the square of  $2 \times 2$  matrix explicitly as a map from  $\mathbb{R}^4$  to itself and compute its Jacobian matrix. Then compare your computation with Example 8.1.5.

**Exercise 8.18.** Express the inverse of  $2 \times 2$  matrix explicitly as a map from an open subset of  $\mathbb{R}^4$  to itself and compute its Jacobian matrix at the identity matrix. Then compare your computation with Exercise 8.13.

## Directional Derivative

Suppose a linear function  $a + l(\Delta \vec{x})$  approximates  $f(\vec{x})$  near  $\vec{x}_0$ . Then for any straight line  $\vec{x}_0 + t\vec{v}$  passing through  $\vec{x}_0$ , the restriction  $a + l(t\vec{v}) = a + l(\vec{v})t$  is a linear function of  $t$  that approximates the restriction  $f(\vec{x}_0 + t\vec{v})$ . This is equivalent to the existence of the derivative of the single variable function  $f(\vec{x}_0 + t\vec{v})$  at  $t = 0$ .

**Definition 8.1.2.** Suppose  $f$  is defined near  $\vec{x}_0$  and  $\vec{v}$  is a vector of unit Euclidean

length. The *directional derivative* of  $f$  at  $\vec{x}_0$  along  $\vec{v}$  is

$$D_{\vec{v}}f(\vec{x}_0) = \left. \frac{d}{dt} \right|_{t=0} f(\vec{x}_0 + t\vec{v}) = \lim_{t \rightarrow 0} \frac{f(\vec{x}_0 + t\vec{v}) - f(\vec{x}_0)}{t}.$$

The partial derivatives are the derivatives along the standard basis vectors  $\vec{e}_i$ . We assume that  $\vec{v}$  has unit length because we wish to forget about the *length* when considering the *direction*. For general  $\vec{v}$ , we need to modify the vector to have unit length. This means that the derivative of  $f$  in the direction of  $\vec{v}$  is  $D_{\frac{\vec{v}}{\|\vec{v}\|_2}}f$ .

If  $f$  is differentiable at  $\vec{x}_0$ , then by the remark before the definition,  $f$  has derivative along any direction, with (note that  $l = f'(\vec{x}_0)$ )

$$D_{\vec{v}}f(\vec{x}_0) = f'(\vec{x}_0)(\vec{v}) = \nabla f(\vec{x}_0) \cdot \vec{v}.$$

**Example 8.1.10.** The function  $f = x^2 + y^2 + z^2$  is differentiable with gradient  $\nabla f = (2x, 2y, 2z)$ . The derivative at  $(1, 1, 1)$  in the direction  $(2, 1, 2)$  is

$$D_{\frac{1}{3}(2,1,2)}f(1,1,1) = \nabla f(1,1,1) \cdot \frac{1}{3}(2,1,2) = \frac{1}{3}(2,2,2) \cdot (2,1,2) = \frac{10}{3}.$$

**Example 8.1.11.** Suppose  $f$  is differentiable, and the derivative of  $f$  in the directions  $(1, 1)$  and  $(1, -1)$  are respectively 2 and  $-3$ . Then

$$\begin{aligned} D_{\frac{1}{\sqrt{2}}(1,1)}f &= (f_x, f_y) \cdot \frac{1}{\sqrt{2}}(1,1) = \frac{f_x + f_y}{\sqrt{2}} = 2, \\ D_{\frac{1}{\sqrt{2}}(1,-1)}f &= (f_x, f_y) \cdot \frac{1}{\sqrt{2}}(1,-1) = \frac{f_x - f_y}{\sqrt{2}} = -3. \end{aligned}$$

This gives the partial derivatives  $f_x = -\frac{1}{\sqrt{2}}$  and  $f_y = \frac{5}{\sqrt{2}}$ .

Note that if  $f$  is not assumed differentiable, then the existence of the derivatives in the directions  $(1, 1)$  and  $(1, -1)$  does not necessarily imply the partial derivatives.

**Example 8.1.12.** Consider the function  $f(x, y) = \frac{x^2y}{x^2 + y^2}$  for  $(x, y) \neq (0, 0)$  and  $f(0, 0) = 0$ . It has partial derivatives  $f_x(0, 0) = f_y(0, 0) = 0$ . Therefore if the function is differentiable at  $(0, 0)$ , then the directional derivative will be  $D_{\vec{v}}f(0, 0) = (0, 0) \cdot \vec{v} = 0$ . On the other hand, for  $\vec{v} = (a, b)$ , by the definition of directional derivative, we have

$$D_{\vec{v}}f = \lim_{t \rightarrow 0} \frac{1}{t} \frac{t^3 a^2 b}{t^2(a^2 + b^2)} = a^2 b.$$

Since this is not always zero, we conclude that the function is not differentiable at  $(0, 0)$ .

**Example 8.1.13.** The function in Example 6.2.3 has zero directional derivative at  $(0, 0)$  in every direction. So the equality  $D_{\vec{v}}f(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \vec{v}$  is satisfied. However, the function may not even be continuous, let alone differentiable.

**Exercise 8.19.** Suppose  $Df = 1$  in direction  $(1, 2, 2)$ ,  $Df = \sqrt{2}$  in direction  $(0, 1, -1)$ ,  $Df = 3$  in direction  $(0, 0, 1)$ . Find the gradient of  $f$ .

**Exercise 8.20.** Suppose  $f$  is differentiable and  $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$  form an orthonormal basis. Prove that

$$\nabla f = (D_{\vec{u}_1} f)\vec{u}_1 + (D_{\vec{u}_2} f)\vec{u}_2 + \cdots + (D_{\vec{u}_n} f)\vec{u}_n.$$

### Condition for Differentiability

A differentiable map must have all the partial derivatives. However, Example 8.1.8 shows that the existence of partial derivatives does not necessarily imply the differentiability. In fact, the partial derivatives only give a candidate linear approximation, and further argument is needed to verify that the candidate indeed approximates. The following result shows that, under slightly stronger condition, the verification is automatic.

**Proposition 8.1.3.** *Suppose a map has all the partial derivatives near  $\vec{x}_0$ , and the partial derivatives are continuous at  $\vec{x}_0$ . Then the map is differentiable at  $\vec{x}_0$ .*

A differentiable map  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is *continuously differentiable* if the map  $\vec{x} \mapsto F'(\vec{x}): \mathbb{R}^n \rightarrow \mathbb{R}^{mn}$  is continuous. This is equivalent to the continuity of all the partial derivatives.

*Proof.* We only prove for a two variable function  $f(x, y)$ . The general case is similar.

Suppose the partial derivatives  $f_x(x, y)$  and  $f_y(x, y)$  exist near  $(x_0, y_0)$ . Applying the Mean Value Theorem to  $f(t, y)$  for  $t \in [x_0, x]$  and to  $f(x_0, s)$  for  $s \in [y_0, y]$ , we get

$$\begin{aligned} f(x, y) - f(x_0, y_0) &= (f(x, y) - f(x_0, y)) + (f(x_0, y) - f(x_0, y_0)) \\ &= f_x(c, y)\Delta x + f_y(x_0, d)\Delta y, \end{aligned}$$

for some  $c \in [x_0, x]$  and  $d \in [y_0, y]$ . If  $f_x$  and  $f_y$  are continuous at  $(x_0, y_0)$ , then for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that (the  $L^1$ -norm is used for  $\Delta\vec{x} = (\Delta x, \Delta y)$ )

$$|\Delta x| + |\Delta y| < \delta \implies |f_x(x, y) - f_x(x_0, y_0)| < \epsilon, |f_y(x, y) - f_y(x_0, y_0)| < \epsilon.$$

By  $|c - x_0| \leq |\Delta x|$  and  $|d - y_0| \leq |\Delta y|$ , we get

$$\begin{aligned} |\Delta x| + |\Delta y| < \delta &\implies |c - x_0| + |d - y_0| \leq \delta, |x_0 - x_0| + |d - y_0| \leq \delta \\ &\implies |f_x(c, y) - f_x(x_0, y_0)| < \epsilon, |f_y(x_0, d) - f_y(x_0, y_0)| < \epsilon \\ &\implies |f(x, y) - f(x_0, y_0) - f_x(x_0, y_0)\Delta x - f_y(x_0, y_0)\Delta y| \\ &= |(f_x(c, y) - f_x(x_0, y_0))\Delta x + (f_y(x_0, d) - f_y(x_0, y_0))\Delta y| \\ &\leq \epsilon|\Delta x| + \epsilon|\Delta y|. \end{aligned}$$

This proves that  $f(x, y)$  is differentiable at  $(x_0, y_0)$ . □

**Example 8.1.14.** The function  $f(x, y) = 1 + 2x + xy^2$  in Example 8.1.7 has continuous partial derivatives  $f_x = 2 + y^2$ ,  $f_y = 2xy$ . Therefore the function is differentiable.

The partial derivatives in Example 8.1.9 are continuous. Therefore the map in Example 8.1.2 is differentiable.

**Example 8.1.15.** On the plane, the polar coordinate  $(r, \theta)$  and the cartesian coordinate  $(x, y)$  are related by  $x = r \cos \theta$ ,  $y = r \sin \theta$ . The relation is differentiable because the partial derivatives are continuous. The Jacobian matrix  $\frac{\partial(x, y)}{\partial(r, \theta)} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}$  and the differential  $\begin{pmatrix} dx \\ dy \end{pmatrix} = \frac{\partial(x, y)}{\partial(r, \theta)} \begin{pmatrix} dr \\ d\theta \end{pmatrix} = \begin{pmatrix} \cos \theta dr - r \sin \theta d\theta \\ \sin \theta dr + r \cos \theta d\theta \end{pmatrix}$ .

**Example 8.1.16.** By Proposition 3.1.5, the differentiability of a parameterized curve

$$\phi(t) = (x_1(t), x_2(t), \dots, x_n(t)): (a, b) \rightarrow \mathbb{R}^n$$

is equivalent to the existence of the derivatives  $x'_i$ . Here the derivatives are not required to be continuous. The Jacobian matrix is the vertical version of the *tangent vector*

$$\phi' = (x'_1, x'_2, \dots, x'_n).$$

The derivative linear transform is  $\phi': u \in \mathbb{R} \mapsto u\phi' \in \mathbb{R}^n$ , the multiplication by the tangent vector. Note that  $\phi'$  is used to denote the linear transform as well as the tangent vector. The vector is the value of the linear transform at  $u = 1$ .

The definition of parameterized curve allows the constant curve and allows the tangent vector to become zero. It also allows continuously differentiable curves to appear to have sharp corners, such as this example

$$\phi(t) = \begin{cases} (t^2, 0), & \text{if } t \leq 0, \\ (0, t^2), & \text{if } t > 0. \end{cases}$$

To avoid such odd situations, we say that a differentiable parameterized curve is *regular* if the tangent vector  $\phi'$  is never zero. In Example 8.4.5, we will explain that a continuously differentiable regular curve can be locally expressed as all coordinates being differentiable functions of some coordinate.

**Example 8.1.17.** The parameterized sphere (6.2.1) and the parameterized torus (6.2.2) are differentiable surfaces in  $\mathbb{R}^3$ . In general, if a parameterized surface  $\sigma(u, v): \mathbb{R}^2 \rightarrow \mathbb{R}^n$  is differentiable, then the tangent vectors  $\sigma_u$  and  $\sigma_v$  span the *tangent plane*  $T_{(u_0, v_0)}S$ . As a linear map, the derivative  $\sigma'$  is  $(s, t) \mapsto s\sigma_u + t\sigma_v$ .

Similar to parameterized curves, we say that  $\sigma$  is *regular* if the tangent vectors  $\sigma_u$  and  $\sigma_v$  are always linearly independent (so that the tangent plane  $T_{(u_0, v_0)}S$  is indeed two dimensional). In Example 8.4.6, we will explain that, in case  $\sigma_u$  and  $\sigma_v$  are continuous, this is equivalent to the possibility of expressing all coordinates as continuously differentiable functions of two coordinates.

**Exercise 8.21.** Compute the Jacobian matrix and the differential. Explain why the maps are differentiable.

1.  $r = \sqrt{x^2 + y^2}$ ,  $\theta = \arctan \frac{y}{x}$ .
2.  $u_1 = x_1 + x_2 + x_3$ ,  $u_2 = x_1x_2 + x_2x_3 + x_3x_1$ ,  $u_3 = x_1x_2x_3$ .
3.  $x = r \sin \phi \cos \theta$ ,  $y = r \sin \phi \sin \theta$ ,  $z = r \cos \phi$ .

**Exercise 8.22.** Suppose a map  $F(\vec{x})$  defined for  $\vec{x}$  near  $\vec{0}$  satisfies  $\|F(\vec{x})\| \leq \|\vec{x}\|^p$  for some  $p > 1$ . Show that  $F(\vec{x})$  is differentiable at  $\vec{0}$ . Therefore the sufficient condition in Proposition 8.1.3 is not a necessary condition for the differentiability.

**Exercise 8.23.** Construct a function that is differentiable everywhere but the partial derivatives are not continuous at some point. What does the example tell you about Proposition 8.1.3?

**Exercise 8.24.** Suppose  $f_x(x_0, y_0)$  exists. Suppose  $f_y(x, y)$  exists for  $(x, y)$  near  $(x_0, y_0)$ , and is continuous at  $(x_0, y_0)$ . Prove that  $f(x, y)$  is differentiable at  $(x_0, y_0)$ . Extend the fact to three or more variables.

**Exercise 8.25.** Suppose  $f(x)$  is a differentiable single variable function. Then by Exercise 6.51, the following function is continuous

$$F(x, y) = \begin{cases} \frac{f(x) - f(y)}{x - y}, & \text{if } x \neq y, \\ f'(x), & \text{if } x = y. \end{cases}$$

Prove the following.

1.  $F$  is differentiable at  $(x_0, x_0)$  if and only if  $f$  has second order derivative at  $x_0$ .
2.  $F'$  is continuous at  $(x_0, x_0)$  if and only if  $f''$  is continuous at  $x_0$ .

## 8.2 Property of Linear Approximation

In Section 3.2, we developed the property of linear approximations as part of a more general theory about approximations. All the discussions about the more general theory can be carried to the multivariable. First we may extend the general definition of approximation (Definition 3.1.1).

**Definition 8.2.1.** A map  $F(\vec{x})$  is approximated near  $\vec{x}_0$  by a map  $P(\vec{x})$  with respect to the base unit function  $u(\vec{x}) \geq 0$ , denoted  $F \sim_u P$ , if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|\vec{x} - \vec{x}_0\| < \delta \implies \|F(\vec{x}) - P(\vec{x})\| \leq \epsilon u(\vec{x}).$$

If we choose  $u(\vec{x}) = 1$  and  $P(\vec{x}) = \vec{a}$ , then we get the definition of continuity at  $\vec{x}_0$ . If we choose  $u(\vec{x}) = \|\vec{x} - \vec{x}_0\|$  and  $P(\vec{x}) = \vec{a} + L(\vec{x} - \vec{x}_0)$ , then we get the definition of differentiability.

The usual properties of general approximations can be extended, including Propositions 3.2.3 to 3.2.5. The extension of Proposition 3.2.2 is more complicated, and will be given by Theorem 8.3.1.

**Exercise 8.26.** Extend Exercises 3.45 to 3.4.

1.  $F \sim_u P$  near  $\vec{x}_0$  implies  $F(\vec{x}_0) = P(\vec{x}_0)$ .
2.  $F \sim_u P$  if and only if  $P \sim_u F$ .
3.  $F \sim_u P$  if and only if  $F - P \sim_u \vec{0}$ .
4. Suppose  $u(\vec{x}) \neq 0$  for  $\vec{x} \neq \vec{x}_0$ . Prove that  $F \sim_u \vec{0}$  near  $\vec{x}_0$  if and only if  $F(\vec{x}_0) = \vec{0}$  and  $\lim_{\vec{x} \rightarrow \vec{x}_0} \frac{F(\vec{x})}{u(\vec{x})} = 0$ .
5. Suppose  $u(\vec{x}) \leq Cv(\vec{x})$  for a constant  $C$ . Prove that  $F \sim_v P$  implies  $F \sim_u P$ .

Exercise 8.27. Extend Proposition 3.2.3: If  $F \sim_u P$  and  $G \sim_u Q$ , then  $F + G \sim_u P + Q$ .

Exercise 8.28. Extend Proposition 3.2.4: Suppose  $B$  is a bilinear map, and  $P, Q, u$  are bounded near  $\vec{x}_0$ . If  $F \sim_u P$  and  $G \sim_u Q$ , then  $B(F, G) \sim_u B(P, Q)$ .

Exercise 8.29. Extend Proposition ???: If  $F \sim_u G$  and  $G \sim_u H$ , then  $F \sim_u H$ .

Exercise 8.30. Extend Proposition 3.2.5: Suppose  $F(\vec{x}) \sim_{u(\vec{x})} P(\vec{x})$  near  $\vec{x}_0$  and  $G(\vec{y}) \sim_{v(\vec{y})} Q(\vec{y})$  near  $\vec{y}_0 = f(\vec{x}_0)$ . Suppose

1.  $P$  is continuous at  $\vec{x}_0$ ,
2.  $u$  is bounded near  $\vec{x}_0$ ,
3.  $v(F(\vec{x})) \leq Au(\vec{x})$  near  $\vec{x}_0$  for a constant  $A$ ,
4.  $|Q(\vec{y}_1) - Q(\vec{y}_2)| \leq B|\vec{y}_1 - \vec{y}_2|$  for  $\vec{y}_1, \vec{y}_2$  near  $\vec{y}_0$ .

Then  $G(F(\vec{x})) \sim_{u(\vec{x})} Q(P(\vec{x}))$ .

## Arithmetic Rule and Leibniz Rule

Given the extension of general approximation theory, the usual rules for computing derivatives can be extended to multivariables.

Suppose  $F, G: \mathbb{R}^n \rightarrow \mathbb{R}^m$  are differentiable at  $\vec{x}_0$ . Then the sum  $F + G: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is also differentiable at  $\vec{x}_0$ , with the derivative linear transform given by

$$(F + G)' = F' + G'.$$

This means that the Jacobian matrix of  $F + G$  is the sum of the Jacobian matrices of  $F$  and  $G$ . In terms of the individual entries of the Jacobian matrix, this means

$$\frac{\partial(f+g)}{\partial x_i} = \frac{\partial f}{\partial x_i} + \frac{\partial g}{\partial x_i}.$$

There are many different versions of “multivariable products”, such that the scalar product, inner product, cross product. All the “products” of two maps  $F: \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$  and  $G: \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$  are given by  $B(F, G): \mathbb{R}^n \rightarrow \mathbb{R}^k$ , for some bilinear map  $B: \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \rightarrow \mathbb{R}^k$ . The general *Leibniz rule* is

$$B(F, G)' = B(F', G) + B(F, G').$$

Here is the specific meaning of the formula at  $\vec{x}_0 \in \mathbb{R}^n$ . For the map  $B(F, G)$  from  $\mathbb{R}^n$  to  $\mathbb{R}^k$ , the derivative  $B(F, G)'(\vec{x}_0)$  is a linear transform between the two vector spaces. Moreover,  $F'(\vec{x}_0)$  is a linear transform that takes  $\vec{v} \in \mathbb{R}^n$  to  $F'(\vec{x}_0)(\vec{v}) \in \mathbb{R}^{m_1}$ , and  $G(\vec{x}_0)$  is a vector in  $\mathbb{R}^{m_2}$ . Therefore  $B(F'(\vec{x}_0), G(\vec{x}_0))$  is a linear transform that takes  $\vec{v} \in \mathbb{R}^n$  to  $B(F'(\vec{x}_0)(\vec{v}), G(\vec{x}_0)) \in \mathbb{R}^k$ . Similarly,  $B(F(\vec{x}_0), G'(\vec{x}_0))$  is also a linear transform from  $\mathbb{R}^n$  to  $\mathbb{R}^k$ . Therefore both sides of the general Leibniz rule are linear transforms from  $\mathbb{R}^n$  to  $\mathbb{R}^k$ .

Each coordinate of  $B$  is a bilinear function. In terms of the individual terms in the bilinear function, the general Leibniz rule means  $\frac{\partial(fg)}{\partial x_i} = \frac{\partial f}{\partial x_i}g + f\frac{\partial g}{\partial x_i}$ .

Exercise 8.31. Prove the general Leibniz rule.



Exercise 8.32. Extend the general Leibniz rule to  $G(F_1, \dots, F_k)$ , for a multilinear map  $G$ .

Exercise 8.33. Suppose  $F, G$  are differentiable at  $\vec{x}_0$ , and  $F(\vec{x}_0) = \vec{0}, G(\vec{x}_0) = \vec{0}$ .

1. Prove that for any bilinear map  $B$ , we have

$$B(F, G) \sim_{\|\Delta\vec{x}\|^2} B(F'(\vec{x}_0)(\Delta\vec{x}), G'(\vec{x}_0)(\Delta\vec{x})).$$

2. Prove that for any quadratic map  $Q$ , we have

$$Q(F) \sim_{\|\Delta\vec{x}\|^2} Q(F'(\vec{x}_0)(\Delta\vec{x})).$$

Moreover, extend to multilinear functions of more variables.

Exercise 8.34. Assuming multilinear maps and functions are differentiable, find the gradients of  $f + g$ ,  $fg$ ,  $F \cdot G$ .

## Chain Rule

Suppose  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable at  $\vec{x}_0$  and  $G: \mathbb{R}^m \rightarrow \mathbb{R}^k$  is differentiable at  $\vec{y}_0 = F(\vec{x}_0)$ . Then the composition  $G \circ F: \mathbb{R}^n \rightarrow \mathbb{R}^k$  is also differentiable at  $\vec{x}_0$ , with the derivative given by the *chain rule*

$$(G \circ F)' = G' \circ F'.$$

The right side is the composition of the linear transforms  $F'(\vec{x}_0): \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $G'(\vec{y}_0) = G'(F(\vec{x}_0)): \mathbb{R}^m \rightarrow \mathbb{R}^k$ . This means that the Jacobian matrix of the composition is the product of the Jacobian matrices of the maps. In terms of the individual entries of the Jacobian matrix, the chain rule means

$$\frac{\partial(g \circ F)}{\partial x_i} = \frac{\partial g}{\partial f_1} \frac{\partial f_1}{\partial x_i} + \frac{\partial g}{\partial f_2} \frac{\partial f_2}{\partial x_i} + \cdots + \frac{\partial g}{\partial f_m} \frac{\partial f_m}{\partial x_i}.$$

Exercise 8.35. Prove the chain rule for multivariable maps.

Exercise 8.36. Discuss the differentiability of a function  $f(\|\vec{x}\|_2)$  of the Euclidean norm.

Exercise 8.37. Suppose  $F, G: \mathbb{R}^n \rightarrow \mathbb{R}^n$  are maps that are inverse to each other. Suppose  $F$  is differentiable at  $\vec{x}_0$  and  $G$  is differentiable at  $\vec{y}_0 = F(\vec{x}_0)$ . Prove that the linear transforms  $F'(\vec{x}_0)$  and  $G'(\vec{y}_0)$  are inverse to each other.

**Example 8.2.1.** This example shows how to calculate the chain rule by the linear relations between the differentials.

Suppose  $u = x^2 + y^2$ ,  $v = xy$ ,  $x = r \cos \theta$ ,  $y = r \sin \theta$ . The relations can be considered as a composition  $(r, \theta) \mapsto (x, y) \mapsto (u, v)$ . The Jacobian matrix of the map  $(r, \theta) \mapsto (x, y)$  can be presented as a linear transform  $(dr, d\theta) \mapsto (dx, dy)$  of the differentials

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \frac{\partial(x, y)}{\partial(r, \theta)} \begin{pmatrix} dr \\ d\theta \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \begin{pmatrix} dr \\ d\theta \end{pmatrix}.$$

The Jacobian matrix of the map  $(x, y) \mapsto (u, v)$  can be presented as a linear transform  $(dx, dy) \mapsto (du, dv)$  of the differentials

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \frac{\partial(u, v)}{\partial(x, y)} \begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} 2x & 2y \\ y & x \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}.$$

The composition of the linear transforms is given by the product of the Jacobian matrices

$$\begin{aligned} \begin{pmatrix} du \\ dv \end{pmatrix} &= \begin{pmatrix} 2x & 2y \\ y & x \end{pmatrix} \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \begin{pmatrix} dr \\ d\theta \end{pmatrix} \\ &= \begin{pmatrix} 2x \cos \theta + 2y \sin \theta & -2xr \sin \theta + 2yr \cos \theta \\ y \cos \theta + x \sin \theta & -yr \sin \theta + xr \cos \theta \end{pmatrix} \begin{pmatrix} dr \\ d\theta \end{pmatrix} \\ &= \begin{pmatrix} 2r & 0 \\ 2r \sin \theta \cos \theta & r^2(\cos^2 \theta - \sin^2 \theta) \end{pmatrix} \begin{pmatrix} dr \\ d\theta \end{pmatrix} \end{aligned}$$

Therefore

$$\frac{\partial u}{\partial r} = 2r, \quad \frac{\partial u}{\partial \theta} = 0, \quad \frac{\partial v}{\partial r} = 2r \sin \theta \cos \theta, \quad \frac{\partial v}{\partial \theta} = r^2(\cos^2 \theta - \sin^2 \theta).$$

**Example 8.2.2.** Suppose  $f(x)$  and  $g(x)$  are differentiable and  $f(x) > 0$ . This example explains the classical computation of the derivative of  $f(x)^{g(x)}$  by applying the idea in Example 8.2.1 to the function  $\phi(u, v) = u^v$ .

Note that  $f(x)^{g(x)}$  is the composition

$$x \mapsto (u, v) = (f(x), g(x)) \mapsto y = \phi(u, v) = u^v.$$

Then we get the linear relations between the differentials

$$dy = \frac{\partial u^v}{\partial u} du + \frac{\partial u^v}{\partial v} dv = vu^{v-1} du + u^v \log u dv, \quad du = f'(x)dx, \quad dv = g'(x)dx.$$

Substituting  $du$  and  $dv$  into  $dy$ , we get

$$dy = vu^{v-1} f' dx + u^v \log u g' dx = (vu^{v-1} f' + u^v \log u g') dx.$$

Therefore

$$\left(f(x)^{g(x)}\right)' = \frac{dy}{dx} = vu^{v-1} f' + u^v \log u g' = f(x)^{g(x)-1} g(x) f'(x) + f(x)^{g(x)} g'(x) \log f(x).$$

**Example 8.2.3.** This example explains the classical Leibniz rule in terms of the derivative of the multiplication function  $\mu(x, y) = xy$  in Example 8.1.1.

Let  $f(t)$  and  $g(t)$  be differentiable. Then the product  $f(t)g(t)$  is the composition

$$t \in \mathbb{R} \mapsto (x, y) = (f(t), g(t)) \in \mathbb{R}^2 \mapsto z = \mu(x, y) = f(t)g(t).$$

The chain rule says that the derivative of  $f(t)g(t)$  is the composition

$$u \in \mathbb{R} \mapsto (uf'(t), ug'(t)) \in \mathbb{R}^2 \mapsto yuf'(t) + xug'(t) = u(g(t)f'(t) + f(t)g'(t)).$$

Taking the value at  $u = 1$ , we get the usual Leibniz rule  $(f(t)g(t))' = f'(t)g(t) + f(t)g'(t)$ .

We may also derive the classical Leibniz rule in the style of Example 8.2.1, by starting with  $d\mu = ydx + xdy$ .

**Exercise 8.38.** Suppose  $f(x, y)$  is a differentiable function. If  $f(t, t^2) = 1$ ,  $f_x(t, t^2) = t$ , find  $f_y(t, t^2)$ .

**Exercise 8.39.** The curves  $\phi(t) = (t, t^2)$  and  $\psi(t) = (t^2, t)$  intersect at  $(1, 1)$ . Suppose the derivatives of a differentiable function  $f(x, y)$  along the two curves are respectively 2 and 3 at  $(1, 1)$ , what is the gradient of  $f$  at  $(1, 1)$ ?

**Exercise 8.40.** For differentiable  $f$ , express the gradient of  $f(x, y)$  in terms of the partial derivatives  $f_r$ ,  $f_\theta$  and the directions  $\vec{e}_r = \frac{\vec{x}_r}{\|\vec{x}_r\|_2} = (\cos \theta, \sin \theta)$ ,  $\vec{e}_\theta = \frac{\vec{x}_\theta}{\|\vec{x}_\theta\|_2} = (-\sin \theta, \cos \theta)$  in the polar coordinates. Exercise 8.113 is a vast generalization.

**Exercise 8.41.** Prove that a differentiable function  $f(x, y)$  depends only on the angle in the polar coordinate if and only if  $xf_x + yf_y = 0$ . Can you make a similar claim for a differentiable function that depends only on the length.

**Exercise 8.42.** Suppose differentiable functions  $f$  and  $g$  satisfy  $f(u, v, w) = g(x, y, z)$  for  $u = \sqrt{yz}$ ,  $v = \sqrt{zx}$ ,  $w = \sqrt{xy}$ . Prove that  $uf_u + vf_v + wf_w = xg_x + yg_y + zg_z$ . Exercise 8.137 is a vast generalization.

**Exercise 8.43.** Find the partial differential equation characterizing differentiable functions  $f(x, y)$  of the form  $f(x, y) = h(xy)$ . What about functions of the form  $f(x, y) = h\left(\frac{y}{x}\right)$ ?

**Exercise 8.44.** Suppose  $f(\vec{x})$  is a multivariable differentiable function and  $g(t)$  is a single variable differentiable function. What is the gradient of  $g(f(\vec{x}))$ ? Extend your answer to the composition  $g(F(\vec{x}))$  of a multivariable differentiable map  $F$  and a multivariable differentiable function  $g$ .

**Example 8.2.4.** Example 8.1.6 gives the derivative  $\det'(I)(H) = \text{tr} H$  of the determinant function at the identity matrix  $I$ . Although the idea can be used to calculate the derivative at the other matrices, the formula obtained is rather complicated (see Exercise 8.11). For invertible  $A$ , we may use the chain rule to find another formula for  $\det'(A)$ .

The determinant is the composition

$$X \in \mathbb{R}^{n^2} \mapsto Y = A^{-1}X \in \mathbb{R}^{n^2} \mapsto y = \det Y \in \mathbb{R} \mapsto x = (\det A)y \in \mathbb{R}.$$

Starting from  $X_0 = A$  on the left, we get corresponding  $Y_0 = I$ ,  $y_0 = 1$ ,  $x_0 = \det A$ . The corresponding composition of linear approximations at these locations is

$$dX \mapsto dY = A^{-1}dX \mapsto dy = \text{tr} dY \mapsto dx = (\det A)dy.$$

Thus we get  $dx = (\det A)\text{tr}(A^{-1}dX)$ . In other words, the derivative of the determinant function at  $A$  is  $\det'(A)(H) = (\det A)\text{tr}(A^{-1}H)$ .

**Exercise 8.45.** For any invertible matrix  $A$ , the inverse map is the composition of the following three maps

$$X \mapsto A^{-1}X, \quad X \mapsto X^{-1}, \quad X \mapsto XA^{-1}.$$

Use this and Exercise 8.13 to find the derivative of the inverse matrix map at  $A$ .

**Example 8.2.5.** The restriction of a function  $f$  on a parameterized curve  $\phi(t): (a, b) \rightarrow \mathbb{R}^n$  is the composition  $f(\phi(t))$ . If both  $f$  and  $\phi$  are differentiable, then the change of  $f$  along  $\phi$  is

$$f(\phi(t))' = f'(\phi(t))(\phi'(t)) = \nabla f(\phi(t)) \cdot \phi'(t) = \|\phi'\|_2 D_{\frac{\phi'}{\|\phi'\|_2}} f.$$

The detailed chain rule formula is

$$\frac{d}{dt} f(x_1(t), \dots, x_n(t)) = \frac{\partial f}{\partial x_1} x_1'(t) + \dots + \frac{\partial f}{\partial x_n} x_n'(t).$$

We emphasise that the chain rule may not hold if  $f$  is not differentiable. For a counterexample, consider  $f(x, y) = \frac{xy}{x^2 + y^2}$  for  $(x, y) \neq (0, 0)$ ,  $f(0, 0) = 0$ , and  $\phi(t) = (t, t^2)$ . On the left, we have

$$f(\phi(t)) = \frac{t}{1 + t^2}, \quad \left. \frac{d}{dt} f(\phi(t)) \right|_{t=0} = 1.$$

On the right, we have  $\phi(0) = (0, 0)$  and

$$f_x(0, 0) = f_y(0, 0) = 0, \quad \phi'(0) = (1, 2t), \quad f_x(0, 0)x'(0) + f_y(0, 0)y'(0) = 0.$$

We see that the chain rule fails.

**Exercise 8.46.** If the chain rule formula in Example 8.2.5 is applied to a straight line  $\phi(t) = \vec{x}_0 + t\vec{v}$ , where  $\vec{v}$  has unit length, then we get the formula  $D_{\vec{v}}f = \nabla f \cdot \vec{v}$  for calculating the directional derivative. Explain that the function in Example 8.1.12 fails the chain rule and is therefore not differentiable.

**Exercise 8.47.** If  $f(\phi(t))$  satisfies the chain rule for all straight lines  $\phi(t)$ , is it true that  $f$  is differentiable?

**Exercise 8.48.** In Examples 8.1.12 and 8.2.5, we saw that the chain rule may not hold for  $F \circ G$  if  $F$  only has partial derivatives but is not differentiable. Can you find a counterexample where  $G$  only has partial derivatives but is not differentiable?

## Mean Value Theorem

Suppose a function  $f(\vec{x})$  is differentiable along the straight line connecting  $\vec{a}$  to  $\vec{b}$

$$\phi(t) = (1 - t)\vec{a} + t\vec{b}, \quad t \in [0, 1].$$

Then the restriction of the function on the straight line is a single variable differentiable function  $g(t) = f((1 - t)\vec{a} + t\vec{b})$  on  $[0, 1]$ , and we have the Mean Value Theorem for the multivariable function  $f$

$$f(\vec{b}) - f(\vec{a}) = g(1) - g(0) = g'(c)(1 - 0) = f'(\vec{c})(\vec{b} - \vec{a}) = \nabla f(\vec{c}) \cdot (\vec{b} - \vec{a}), \quad c \in (0, 1),$$

where the second equality is the Mean Value Theorem of single variable function and the last two equalities follow from Example 8.2.5.

For a multivariable map, the Mean Value Theorem may be applied to each coordinate function. However, the choice of  $\vec{c}$  may be different for different coordinate. The following is a more unified extension.

**Proposition 8.2.2.** *Suppose a map  $F$  is differentiable along the straight line connecting  $\vec{a}$  and  $\vec{b}$ . Then there is  $\vec{c}$  on the straight line, such that*

$$\|F(\vec{b}) - F(\vec{a})\|_2 \leq \|F'(\vec{c})\| \|\vec{b} - \vec{a}\|_2.$$

where the norm  $\|F'(\vec{c})\|$  is induced from the Euclidean norms.

*Proof.* For any fixed vector  $\vec{u}$ , the function  $f(\vec{x}) = F(\vec{x}) \cdot \vec{u}$  is the composition  $\vec{x} \mapsto \vec{y} = F(\vec{x}) \mapsto z = \vec{y} \cdot \vec{u}$ . Since the second map is linear, the derivative  $f'(\vec{c})$  is the composition

$$d\vec{x} \mapsto d\vec{y} = F'(\vec{c})(d\vec{x}) \mapsto dz = d\vec{y} \cdot \vec{u} = F'(\vec{c})(d\vec{x}) \cdot \vec{u}.$$

By the discussion before the proposition, we have

$$(F(\vec{b}) - F(\vec{a})) \cdot \vec{u} = f(\vec{b}) - f(\vec{a}) = f'(\vec{c})(\vec{b} - \vec{a}) = F'(\vec{c})(\vec{b} - \vec{a}) \cdot \vec{u}.$$

By Schwarz's inequality, we have

$$|(F(\vec{b}) - F(\vec{a})) \cdot \vec{u}| \leq \|F'(\vec{c})(\vec{b} - \vec{a})\|_2 \|\vec{u}\|_2 \leq \|F'(\vec{c})\| \|\vec{b} - \vec{a}\|_2 \|\vec{u}\|_2.$$

If we take  $\vec{u} = F(\vec{b}) - F(\vec{a})$  at the beginning, then we get  $\|F(\vec{b}) - F(\vec{a})\|_2 \leq \|F'(\vec{c})\| \|\vec{b} - \vec{a}\|_2$ .  $\square$

**Exercise 8.49.** What can you say about Proposition 8.2.2 if the norms are not Euclidean?

**Exercise 8.50.** Suppose a differentiable map on a path connected open subset has zero derivative everywhere. Prove that the map is a constant map. This extends Proposition 3.3.4. What if only some partial derivatives are constantly zero?

## 8.3 Inverse and Implicit Differentiations

Two primary problems about multivariable maps are the invertibility of maps (such as the change of variables between the cartesian and the polar coordinates), and solving systems of equations. The problems can be solved if the corresponding linear approximation problems can be solved.

### Inverse Differentiation

Suppose a single variable function  $f(x)$  has continuous derivative near  $x_0$ . If  $f'(x_0) \neq 0$ , then  $f'(x)$  is either positive for all  $x$  near  $x_0$  or negative for all  $x$  near  $x_0$ . Thus  $f(x)$  is strictly monotone and is therefore invertible near  $x_0$ . Moreover, Theorem 3.2.2 says that the inverse function is also differentiable at  $x_0$ , with the expected linear approximation.

The multivariable extension is the following.

**Theorem 8.3.1 (Inverse Function Theorem).** *Suppose  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is continuously differentiable near  $\vec{x}_0$ . Suppose the derivative  $F'(\vec{x}_0)$  is an invertible linear map.*

Then there is an open subset  $U$  around  $\vec{x}_0$ , such that  $F(U)$  is also open,  $F: U \rightarrow F(U)$  is invertible,  $F^{-1}: F(U) \rightarrow U$  is differentiable, and  $(F^{-1})'(\vec{y}) = (F'(\vec{x}))^{-1}$  when  $\vec{y} = F(\vec{x})$ .

The theorem basically says that if the linear approximation of a map is invertible, then the map is also invertible, at least locally. Exercise 3.41 shows that the continuity assumption cannot be dropped from the theorem.

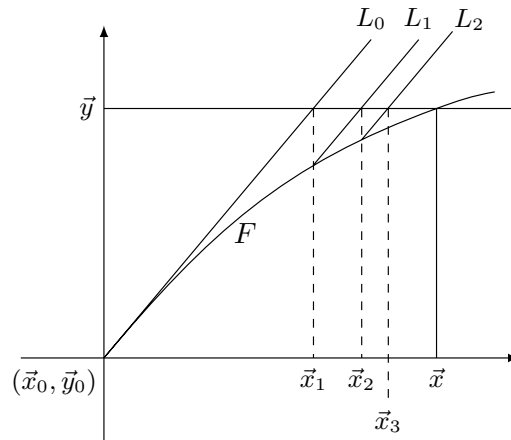
The inverse function  $F^{-1}$  is also differentiable, and should be approximated by the inverse linear transform  $(F')^{-1}$ . This is the last conclusion of the theorem and can be compared with Theorem 3.2.2.

*Proof.* For any  $\vec{y}$  near  $\vec{y}_0 = F(\vec{x}_0)$ , we wish to find  $\vec{x}$  near  $\vec{x}_0$  satisfying  $F(\vec{x}) = \vec{y}$ . To solve the problem, we approximate the map  $F(\vec{x})$  by the linear map  $L_0(\vec{x}) = F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x} - \vec{x}_0)$  and solve the similar approximate linear equation  $L_0(\vec{x}) = \vec{y}$ . The solution  $\vec{x}_1$  of  $L_0(\vec{x}) = \vec{y}$  satisfies

$$\vec{y} = F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x}_1 - \vec{x}_0) = \vec{y}_0 + F'(\vec{x}_0)(\vec{x}_1 - \vec{x}_0). \quad (8.3.1)$$

Although not exactly equal to the solution  $\vec{x}$  that we are looking for,  $\vec{x}_1$  is often closer to  $\vec{x}$  than  $\vec{x}_0$ . See Figure 8.3.1. So we repeat the process by using  $L_1(\vec{x}) = F(\vec{x}_1) + F'(\vec{x}_0)(\vec{x} - \vec{x}_1)$  to approximate  $F$  near  $\vec{x}_1$  and solve the similar approximate linear equation  $L_1(\vec{x}) = \vec{y}$  to get solution  $\vec{x}_2$ . Continuing the process with the linear approximation  $L_2(\vec{x}) = F(\vec{x}_2) + F'(\vec{x}_0)(\vec{x} - \vec{x}_2)$  of  $F$  near  $\vec{x}_2$  and so on, we get a sequence  $\vec{x}_k$  inductively defined by

$$\vec{y} = L_k(\vec{x}_{k+1}) = F(\vec{x}_k) + F'(\vec{x}_0)(\vec{x}_{k+1} - \vec{x}_k). \quad (8.3.2)$$



**Figure 8.3.1.** Find  $\vec{x}$  satisfying  $F(\vec{x}) = \vec{y}$ .

To prove the expectation that the sequence  $\vec{x}_k$  converges to a solution  $\vec{x}$  of the equation  $F(\vec{x}) = \vec{y}$ , we use the approximation of  $F$  near  $\vec{x}_0$

$$F(\vec{x}) = F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x} - \vec{x}_0) + R(\vec{x}), \quad R(\vec{x}) = o(\|\vec{x} - \vec{x}_0\|), \quad (8.3.3)$$

to substitute  $F(\vec{x}_k)$  in (8.3.2). We get

$$\begin{aligned}\vec{y} &= F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x}_k - \vec{x}_0) + R(\vec{x}_k) + F'(\vec{x}_0)(\vec{x}_{k+1} - \vec{x}_k) \\ &= F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x}_{k+1} - \vec{x}_0) + R(\vec{x}_k).\end{aligned}\quad (8.3.4)$$

With  $k - 1$  in place of  $k$ , the relation becomes

$$\vec{y} = F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x}_k - \vec{x}_0) + R(\vec{x}_{k-1}).$$

Taking the difference between this and (8.3.4), we get

$$F'(\vec{x}_0)(\vec{x}_{k+1} - \vec{x}_k) + R(\vec{x}_k) - R(\vec{x}_{k-1}) = \vec{0}. \quad (8.3.5)$$

This will give us a relation between  $\vec{x}_{k+1} - \vec{x}_k$  and  $\vec{x}_k - \vec{x}_{k-1}$ .

By the continuity of  $F'$  at  $\vec{x}_0$ , for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $F$  is defined on the ball  $B(\vec{x}_0, \delta)$  and

$$\|\vec{x} - \vec{x}_0\| < \delta \implies \|R'(\vec{x})\| = \|F'(\vec{x}) - F'(\vec{x}_0)\| < \epsilon. \quad (8.3.6)$$

If we know  $\vec{x}_k, \vec{x}_{k-1} \in B(\vec{x}_0, \delta)$ , then the straight line connecting  $\vec{x}_k$  and  $\vec{x}_{k-1}$  still lies in  $B(\vec{x}_0, \delta)$ . By (8.3.5), (8.3.6) and Proposition 8.2.2 (from now on, the norms in the proof are the Euclidean norms),

$$\begin{aligned}\|\vec{x}_{k+1} - \vec{x}_k\| &= \|F'(\vec{x}_0)^{-1}(R(\vec{x}_k) - R(\vec{x}_{k-1}))\| \\ &\leq \|F'(\vec{x}_0)^{-1}\| \|R(\vec{x}_k) - R(\vec{x}_{k-1})\| \\ &\leq \epsilon \|F'(\vec{x}_0)^{-1}\| \|\vec{x}_k - \vec{x}_{k-1}\|.\end{aligned}\quad (8.3.7)$$

If we fix some  $0 < \alpha < 1$  and assume  $\epsilon < \frac{\alpha}{\|F'(\vec{x}_0)^{-1}\|}$  at the beginning, then we get  $\|\vec{x}_{k+1} - \vec{x}_k\| \leq \alpha \|\vec{x}_k - \vec{x}_{k-1}\|$ . Therefore, if we know  $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k \in B(\vec{x}_0, \delta)$ , then

$$\|\vec{x}_{i+1} - \vec{x}_i\| \leq \alpha^i \|\vec{x}_1 - \vec{x}_0\|, \quad 0 \leq i \leq k,$$

so that for any  $0 \leq j < i \leq k + 1$ , we have

$$\|\vec{x}_i - \vec{x}_j\| \leq (\alpha^j + \alpha^{j+1} + \dots + \alpha^{i-1}) \|\vec{x}_1 - \vec{x}_0\| < \frac{\alpha^j}{1 - \alpha} \|\vec{x}_1 - \vec{x}_0\|. \quad (8.3.8)$$

How do we know  $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k \in B(\vec{x}_0, \delta)$ , so that the estimations (8.3.7) and (8.3.8) hold? The estimation (8.3.8) tells us that if  $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k \in B(\vec{x}_0, \delta)$ , then

$$\|\vec{x}_{k+1} - \vec{x}_0\| < \frac{1}{1 - \alpha} \|\vec{x}_1 - \vec{x}_0\| \leq \frac{1}{1 - \alpha} \|F'(\vec{x}_0)^{-1}\| \|\vec{y} - \vec{y}_0\|,$$

where (8.3.1) is used in the second inequality. The right side will be  $< \delta$  if we assume  $\vec{y}$  satisfies

$$\|\vec{y} - \vec{y}_0\| < \frac{(1 - \alpha)\delta}{\|F'(\vec{x}_0)^{-1}\|}. \quad (8.3.9)$$

Under the assumption, we still have  $\vec{x}_{k+1} \in B(\vec{x}_0, \delta)$  and the inductive estimation can continue.

In summary, we should make the following rigorous setup at the very beginning of the proof: Assume  $0 < \alpha < 1$  is fixed,  $0 < \epsilon < \frac{\alpha}{\|F'(\vec{x}_0)^{-1}\|}$  is given, and  $\delta > 0$  is found so that (8.3.6) holds. Then for any  $\vec{y}$  satisfying (8.3.9), we may inductively construct a sequence  $\vec{x}_k$  by the formula (8.3.2). The sequence satisfies (8.3.8). Therefore it is a Cauchy sequence and converges to some  $\vec{x}$ . Taking the limit of (8.3.2), we get  $\vec{y} = F(\vec{x})$ , which means that  $\vec{x}$  is a solution we are looking for. Moreover, the solution  $\vec{x}$  satisfies

$$\|\vec{x} - \vec{x}_0\| = \lim \|\vec{x}_{k+1} - \vec{x}_0\| \leq \frac{1}{1 - \alpha} \|F'(\vec{x}_0)^{-1}\| \|\vec{y} - \vec{y}_0\| < \delta.$$

Geometrically, this means

$$F(B(\vec{x}_0, \delta)) \supset B\left(\vec{y}_0, \frac{(1 - \alpha)\delta}{\|F'(\vec{x}_0)^{-1}\|}\right). \quad (8.3.10)$$

Note that so far we have only used the fact that  $F'$  is continuous at  $\vec{x}_0$  and  $F'(\vec{x}_0)$  is invertible.

Since  $F'$  is continuous and  $F'(\vec{x}_0)$  is invertible, by choosing  $\delta$  even smaller, we may further assume that  $F'(\vec{x})$  is actually invertible for any  $\vec{x} \in U = B(\vec{x}_0, \delta)$ . We claim that the image  $F(U)$  is open. A point in  $F(U)$  is of the form  $F(\vec{x})$  for some  $\vec{x} \in U$ . With  $\vec{x}$  playing the role of  $\vec{x}_0$  in the earlier argument, we may find  $\alpha_x$ ,  $\epsilon_x$  and sufficiently small  $\delta_x$ , with  $B(\vec{x}, \delta_x) \subset U$ . Then for  $\delta'_x = \frac{(1 - \alpha_x)\delta_x}{\|F'(\vec{x})^{-1}\|}$ , the inclusion (8.3.10) becomes

$$B(F(\vec{x}), \delta'_x) \subset F(B(\vec{x}, \delta_x)) \subset F(U).$$

Therefore  $F(U)$  contains a ball around  $F(\vec{x})$ . This proves that  $F(U)$  is open.

Next we prove that  $F$  is one-to-one on the ball  $U$ , so that  $U$  can be used as the open subset in the statement of the theorem. By the approximation (8.3.3) and the estimation (8.3.6), for any  $\vec{x}, \vec{x}' \in U = B(\vec{x}_0, \delta)$ ,

$$\begin{aligned} \|F(\vec{x}) - F(\vec{x}')\| &= \|F'(\vec{x}_0)(\vec{x} - \vec{x}') + R(\vec{x}) - R(\vec{x}')\| \\ &\geq \|F'(\vec{x}_0)(\vec{x} - \vec{x}')\| - \epsilon \|\vec{x} - \vec{x}'\| \\ &\geq \left( \frac{1}{\|F'(\vec{x}_0)^{-1}\|} - \epsilon \right) \|\vec{x} - \vec{x}'\| \\ &\geq \frac{1 - \alpha}{\|F'(\vec{x}_0)^{-1}\|} \|\vec{x} - \vec{x}'\|. \end{aligned} \quad (8.3.11)$$

Since  $\alpha < 1$ ,  $\vec{x} \neq \vec{x}'$  implies  $F(\vec{x}) \neq F(\vec{x}')$ .

It remains to prove the differentiability of the inverse map. By (8.3.3), for  $\vec{y} \in F(U)$  and  $\vec{x} = F^{-1}(\vec{y})$ , we have

$$\vec{x} = \vec{x}_0 + F'(\vec{x}_0)^{-1}(\vec{y} - \vec{y}_0) - F'(\vec{x}_0)^{-1}R(\vec{x}).$$



Then by  $R(\vec{x}_0) = \vec{0}$ ,  $\|R'(\vec{x})\| < \epsilon$  along the line connecting  $\vec{x}_0$  and  $\vec{x}$ , and Proposition 8.2.2, we get

$$\begin{aligned}\|F'(\vec{x}_0)^{-1}R(\vec{x})\| &\leq \|F'(\vec{x}_0)^{-1}\|\|R(\vec{x}) - R(\vec{x}_0)\| \\ &\leq \epsilon\|F'(\vec{x}_0)^{-1}\|\|\vec{x} - \vec{x}_0\| \\ &\leq \frac{\epsilon\|F'(\vec{x}_0)^{-1}\|^2}{1 - \alpha}\|\vec{y} - \vec{y}_0\|,\end{aligned}$$

where the last inequality makes use of (8.3.11). The estimation shows that  $\vec{x}_0 + F'(\vec{x}_0)^{-1}(\vec{y} - \vec{y}_0)$  is a linear approximation of  $F^{-1}(\vec{y})$ . Therefore the inverse map is differentiable at  $\vec{x}_0$ , with  $(F^{-1})'(\vec{y}_0) = F'(\vec{x}_0)^{-1}$ .  $\square$

Note that most of the proof only makes use of the assumption that the derivative  $F'$  is continuous at  $\vec{x}_0$ . The assumption implies that the image of a ball around  $\vec{x}_0$  contains a ball around  $F(\vec{x}_0)$ . Based on this fact, the continuity of the derivative everywhere is (only) further used to conclude that the image of open subsets must be open.

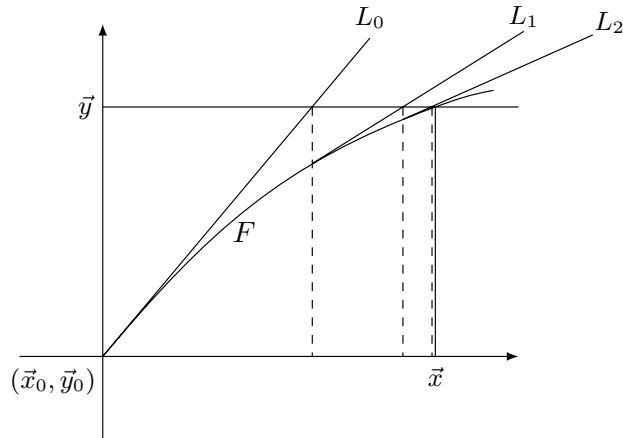
The proof includes a method for computing the inverse function. In fact, it appears to be more efficient to use

$$L_n(\vec{x}) = F(\vec{x}_k) + F'(\vec{x}_k)(\vec{x} - \vec{x}_k)$$

to approximate  $F$  at  $\vec{x}_k$  (with  $F'$  at  $\vec{x}_k$  instead of at  $\vec{x}_0$ ). The method is quite effective in case the dimension is small (so that  $F'(\vec{x}_k)^{-1}$  is easier to compute). In particular, for a single variable function  $f(x)$ , the solution to  $f(x) = 0$  can be found by starting from  $x_0$  and successively constructing

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

This is Newton's method.



**Figure 8.3.2.** *Newton's method.*

**Example 8.3.1.** In Example 8.1.15, we computed the differential of the cartesian coordinate in terms of the polar coordinate. The Jacobian matrix is invertible away from the origin, so that the polar coordinate can also be expressed locally in terms of the cartesian coordinate. In fact, the map  $(r, \theta) \rightarrow (x, y)$  is invertible for  $(r, \theta) \in (0, +\infty) \times (a, a + 2\pi)$ . By solving the system  $dx = \cos \theta dr - r \sin \theta d\theta$ ,  $dy = \sin \theta dr + r \cos \theta d\theta$ , we get

$$dr = \cos \theta dx + \sin \theta dy, \quad d\theta = -r^{-1} \sin \theta dx + r^{-1} \cos \theta dy.$$

By the Inverse Function Theorem, the coefficients form the Jacobian matrix

$$\frac{\partial(r, \theta)}{\partial(x, y)} = \begin{pmatrix} \cos \theta & \sin \theta \\ -r^{-1} \sin \theta & r^{-1} \cos \theta \end{pmatrix}.$$

**Exercise 8.51.** Use the differential in the third part of Exercise 8.21 to find the differential of the change from the cartesian coordinate  $(x, y, z)$  to the spherical coordinate  $(r, \phi, \theta)$ .

**Exercise 8.52.** Suppose  $x = e^u + u \cos v$ ,  $y = e^u + u \sin v$ . Find the places where  $u$  and  $v$  can be locally expressed as differentiable functions of  $x$  and  $y$  and then compute  $\frac{\partial(u, v)}{\partial(x, y)}$ .

**Exercise 8.53.** Find the places where  $z$  can be locally expressed as a function of  $x$  and  $y$  and then compute  $z_x$  and  $z_y$ .

1.  $x = s + t$ ,  $y = s^2 + t^2$ ,  $z = s^3 + t^3$ .
2.  $x = e^{s+t}$ ,  $y = e^{s-t}$ ,  $z = st$ .

**Exercise 8.54.** Change the partial differential equation  $(x + y)u_x - (x - y)u_y = 0$  to an equation with respect to the polar coordinate  $(r, \theta)$ .

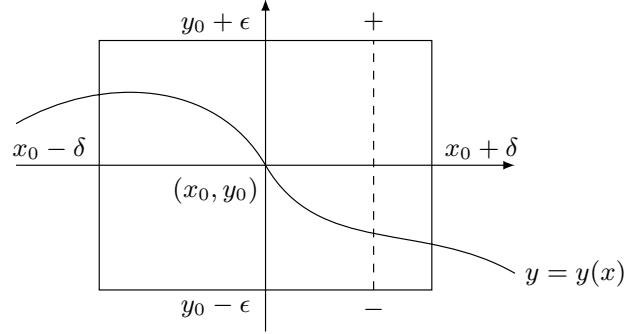
**Exercise 8.55.** Consider the square map  $X \mapsto X^2$  of  $n \times n$  matrices. Find the condition for the square map to have continuously differentiable inverse near a diagonalizable matrix. Moreover, find the derivative of the inverse map (the “square root” of a matrix).

## Implicit Differentiation

Suppose a continuous function  $f(x, y)$  has continuous partial derivative  $f_y$  near  $(x_0, y_0)$  and satisfies  $f(x_0, y_0) = 0$ ,  $f_y(x_0, y_0) \neq 0$ . If  $f_y(x_0, y_0) > 0$ , then  $f_y(x, y) > 0$  for  $(x, y)$  near  $(x_0, y_0)$ . Therefore for any fixed  $x$ ,  $f(x, y)$  is strictly increasing in  $y$ . In particular, for some small  $\epsilon > 0$ , we have  $f(x_0, y_0 + \epsilon) > f(x_0, y_0) = 0$  and  $f(x_0, y_0 - \epsilon) < f(x_0, y_0) = 0$ . By the continuity in  $x$ , there is  $\delta > 0$ , such that  $f(x, y_0 + \epsilon) > 0$  and  $f(x, y_0 - \epsilon) < 0$  for any  $x \in (x_0 - \delta, x_0 + \delta)$ . Now for any fixed  $x \in (x_0 - \delta, x_0 + \delta)$ ,  $f(x, y)$  is strictly increasing in  $y$  and has different signs at  $y_0 + \epsilon$  and  $y_0 - \epsilon$ . Therefore there is a unique  $y = y(x) \in (y_0 - \epsilon, y_0 + \epsilon)$  satisfying  $f(x, y) = 0$ . If  $f_y(x, y) < 0$ , the same argument also gives the unique existence of “solution”  $y(x)$ .

We say  $y(x)$  is an *implicit function* of  $x$  because the function is only implicitly given by the equation  $f(x, y) = 0$ .

The argument shows that if  $f(x_0, y_0) = 0$  and  $f_y(x_0, y_0) \neq 0$ , plus some continuity condition, then the equation  $f(x, y) = 0$  can be solved to define a function



**Figure 8.3.3.** *Implicitly defined function.*

$y = y(x)$  near  $(x_0, y_0)$ . If  $f$  is differentiable at  $(x_0, y_0)$ , then the equation  $f(x, y) = 0$  is approximated by the linear equation

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) = 0.$$

The assumption  $f_y(x_0, y_0) \neq 0$  makes it possible to solve the linear equation and get  $y = y_0 - \frac{f_x(x_0, y_0)}{f_y(x_0, y_0)}(x - x_0)$ . So the conclusion is that if the linear approximation implicitly defines a function, then the original equation also implicitly defines a function.

In general, consider a differentiable map  $F: \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{m+n} \rightarrow \mathbb{R}^m$ . We have

$$\begin{aligned} F'(\vec{x}_0, \vec{y}_0)(\vec{u}, \vec{v}) &= F'(\vec{x}_0, \vec{y}_0)(\vec{u}, \vec{0}) + F'(\vec{x}_0, \vec{y}_0)(\vec{0}, \vec{v}) \\ &= F_{\vec{x}}(\vec{x}_0, \vec{y}_0)(\vec{u}) + F_{\vec{y}}(\vec{x}_0, \vec{y}_0)(\vec{v}), \end{aligned}$$

where  $F_{\vec{x}}(\vec{x}_0, \vec{y}_0): \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $F_{\vec{y}}(\vec{x}_0, \vec{y}_0): \mathbb{R}^m \rightarrow \mathbb{R}^m$  are linear transforms and generalize the partial derivatives. In terms of the Jacobian matrix,  $F'$  can be written as the block matrix  $(F_{\vec{x}} \ F_{\vec{y}})$ .

The equation  $F(\vec{x}, \vec{y}) = \vec{0}$  is approximated by the linear equation

$$F_{\vec{x}}(\vec{x}_0, \vec{y}_0)(\vec{x} - \vec{x}_0) + F_{\vec{y}}(\vec{x}_0, \vec{y}_0)(\vec{y} - \vec{y}_0) = \vec{0}. \quad (8.3.12)$$

The linear equation defines  $\vec{y}$  as a linear map of  $\vec{x}$  (in other words, the linear equation has a unique solution  $\vec{x}$  for each  $\vec{y}$ ) if and only if the linear transform  $F_{\vec{y}}(\vec{x}_0, \vec{y}_0)$  is invertible. We expect the condition to be also the condition for the original equation  $F(\vec{x}, \vec{y}) = \vec{0}$  to locally define  $\vec{y}$  as a map of  $\vec{x}$ . Moreover, this map should be linearly approximated by the solution

$$\Delta \vec{y} = -F_{\vec{y}}^{-1} F_{\vec{x}} \Delta \vec{x}$$

to the linear approximation equation (8.3.12), so that  $-F_{\vec{y}}^{-1} F_{\vec{x}}$  should be the derivative of the map.

**Theorem 8.3.2 (Implicit Function Theorem).** Suppose  $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$  is continuously differentiable near  $(\vec{x}_0, \vec{y}_0)$  and satisfies  $F(\vec{x}_0, \vec{y}_0) = \vec{0}$ . Suppose the derivative  $F_{\vec{y}}(\vec{x}_0, \vec{y}_0): \mathbb{R}^m \rightarrow \mathbb{R}^m$  in  $\vec{y}$  is an invertible linear map. Then there is an open subset  $U$  around  $\vec{x}_0$  and a unique map  $G: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ , such that  $G(\vec{x}_0) = \vec{y}_0$  and  $F(\vec{x}, G(\vec{x})) = \vec{0}$ . Moreover,  $G$  is continuously differentiable near  $\vec{x}_0$  and  $G' = -F_{\vec{y}}^{-1} F_{\vec{x}}$ .

*Proof.* The map  $H(\vec{x}, \vec{y}) = (\vec{x}, F(\vec{x}, \vec{y})): \mathbb{R}^{m+n} \rightarrow \mathbb{R}^{m+n}$  has continuous derivative

$$H'(\vec{u}, \vec{v}) = (\vec{u}, F_{\vec{x}}(\vec{u}) + F_{\vec{y}}(\vec{v}))$$

near  $\vec{x}_0$ . The invertibility of  $F_{\vec{y}}(\vec{x}_0, \vec{y}_0)$  implies the invertibility of  $H'(\vec{x}_0, \vec{y}_0)$ . Then by the Inverse Function Theorem,  $H$  has inverse  $H^{-1} = (S, T)$  that is continuously differentiable near  $H(\vec{x}_0, \vec{y}_0) = (\vec{x}_0, \vec{0})$ , where  $S: \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$  and  $T: \mathbb{R}^{m+n} \rightarrow \mathbb{R}^m$ . Since  $HH^{-1} = (S, F(S, T))$  is the identity, we have  $S(\vec{x}, \vec{z}) = \vec{x}$  and  $F(\vec{x}, T(\vec{x}, \vec{z})) = \vec{z}$ . Then

$$\begin{aligned} F(\vec{x}, \vec{y}) = \vec{0} &\iff H(\vec{x}, \vec{y}) = (\vec{x}, \vec{0}) \\ &\iff (\vec{x}, \vec{y}) = H^{-1}(\vec{x}, \vec{0}) = (S(\vec{x}, \vec{0}), T(\vec{x}, \vec{0})) = (\vec{x}, T(\vec{x}, \vec{0})). \end{aligned}$$

Therefore  $\vec{y} = G(\vec{x}) = T(\vec{x}, \vec{0})$  is exactly the solution of  $F(\vec{x}, \vec{y}) = \vec{0}$ .

Finally, we may differentiate the equality  $F(\vec{x}, G(\vec{x})) = \vec{0}$  to get

$$F_{\vec{x}} + F_{\vec{y}}G' = 0.$$

Since  $F_{\vec{y}}$  is invertible, this implies  $G' = -F_{\vec{y}}^{-1}F_{\vec{x}}$ . □

**Example 8.3.2.** The unit sphere  $S^2 \subset \mathbb{R}^3$  is given by the equation  $f(x, y, z) = x^2 + y^2 + z^2 = 1$ . By  $f_z = 2z$  and the Implicit Function Theorem,  $z$  can be expressed as a function of  $(x, y)$  near any place where  $z \neq 0$ . In fact, the expression is  $z = \pm\sqrt{1 - x^2 - y^2}$ , where the sign is the same as the sign of  $z$ . By solving the equation

$$df = 2xdx + 2ydy + 2zdz = 0,$$

we get  $dz = -\frac{x}{z}dx - \frac{y}{z}dy$ . Therefore  $z_x = -\frac{x}{z}$  and  $z_y = -\frac{y}{z}$ .

**Exercise 8.56.** Find the places where the map is implicitly defined and compute the derivatives of the map.

1.  $x^3 + y^3 - 3axy = 0$ , find  $\frac{dy}{dx}$ .
2.  $x^2 + y^2 + z^2 - 4x + 6y - 2z = 0$ , find  $\frac{\partial z}{\partial(x, y)}$  and  $\frac{\partial x}{\partial(y, z)}$ .
3.  $x^2 + y^2 = z^2 + w^2$ ,  $x + y + z + w = 1$ , find  $\frac{\partial(z, w)}{\partial(x, y)}$  and  $\frac{\partial(x, w)}{\partial(y, z)}$ .
4.  $z = f(x + y + z, xyz)$ , find  $\frac{\partial z}{\partial(x, y)}$  and  $\frac{\partial x}{\partial(y, z)}$ .

**Exercise 8.57.** Verify that the implicitly defined function satisfies the partial differential equation.

1.  $f(x - az, y - bz) = 0$ ,  $z = z(x, y)$  satisfies  $az_x + bz_y = 1$ .
2.  $x^2 + y^2 + z^2 = yf\left(\frac{z}{y}\right)$ ,  $z = z(x, y)$  satisfies  $(x^2 - y^2 - z^2)z_x + 2xyz_y = 4xz$ .

**Exercise 8.58.** In solving the equation  $f(x, y) = 0$  for two variable functions, we did not assume anything about the partial derivative in  $x$ . Extend the discussion to the general multivariable case and point out what conclusion in the Implicit Function Theorem may not hold.

## 8.4 Submanifold

### Hypersurface

A *surface* usually means a nice 2-dimensional subset of  $\mathbb{R}^3$ , and is often given as a level  $f(x, y, z) = c$  of a function. For example, if  $f = x^2 + y^2 + z^2$  is the square of the distance to the origin, then the levels of  $f$  (for  $c > 0$ ) are the spheres. If  $f$  is the distance to a circle in  $\mathbb{R}^3$ , then the levels are the tori. In general, if  $f$  is the distance to a curve, then the surface we get is the tube around the curve.

A *hypersurface* is a nice  $(n - 1)$ -dimensional subset of  $\mathbb{R}^n$ , and is often given as a level

$$S_c = \{\vec{x}: f(\vec{x}) = c\}$$

of a function  $f(\vec{x})$  on  $\mathbb{R}^n$ . The unit sphere  $S^{n-1}$  is the level  $f(x) = 1$  of  $f(x) = \vec{x} \cdot \vec{x} = \|\vec{x}\|_2^2$ . The  $(n - 1)$ -dimensional *hyperplanes* orthogonal to  $\vec{a}$  are the levels of a linear functional  $f(\vec{x}) = \vec{a} \cdot \vec{x}$ . On the other hand, a hypersurface in  $\mathbb{R}^2$  is simply a curve.

A function  $f$  does not change along a parameterized curve  $\phi$  if and only if the curve lies in a level hypersurface  $S_c$ . The *tangent space* of  $S_c$  then consists of the tangent vectors of all the curves inside the level

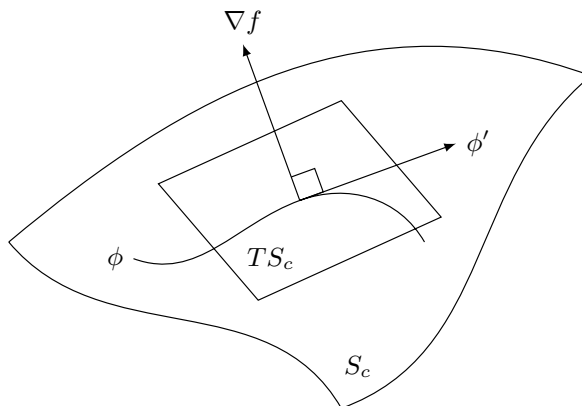
$$\begin{aligned} TS_c &= \{\phi'(t): \phi(t) \in S_c\} \\ &= \{\phi'(t): f(\phi(t)) = c\} \\ &= \{\phi'(t): f(\phi(t))' = \nabla f \cdot \phi'(t) = 0\} \\ &= \{\vec{v}: \nabla f \cdot \vec{v} = 0\}. \end{aligned}$$

This is the vector subspace orthogonal to the gradient. Therefore the unit length vector

$$\vec{n} = \frac{\nabla f}{\|\nabla f\|_2}$$

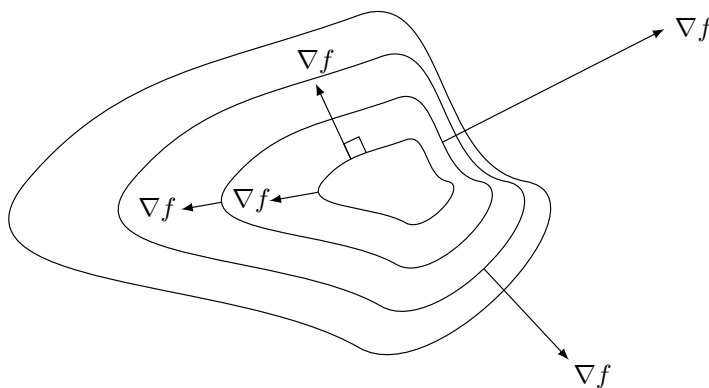
in the direction of the gradient is the *normal vector* of the hypersurface.

On the other hand, the function changes only if one jumps from one level  $S_c$  to a different level  $S_{c'}$ . The jump is measured by the directional derivative  $D_{\vec{v}}f = \nabla f \cdot \vec{v}$ , which is biggest when  $\vec{v}$  is the normal vector  $\vec{n}$ . This means that the change is fastest when one moves orthogonal to the levels. Moreover, the magnitude



**Figure 8.4.1.** *Tangent space of level hypersurface.*

$\|\nabla f\|_2 = D_{\vec{n}}f$  of the gradient measures how fast this fastest jump is. Therefore  $\|\nabla f\|_2$  measures how much the levels are “squeezed” together.



**Figure 8.4.2.** *Longer  $\nabla f$  means the levels are closer to each other.*

**Example 8.4.1.** The graph of a differentiable function  $f(x, y)$  defined on an open subset  $U \subset \mathbb{R}^2$  is a surface

$$S = \{(x, y, f(x, y)) : (x, y) \in U\}.$$

The surface can be considered as a level  $g(x, y, z) = z - f(x, y) = 0$ . The tangent space of  $S$  is orthogonal to  $\nabla g = (-f_x, -f_y, 1)$ . The normal vector of the graph surface is  $\vec{n} = \frac{(-f_x, -f_y, 1)}{\sqrt{f_x^2 + f_y^2 + 1}}$ .

**Example 8.4.2.** The gradient of the function  $\vec{x} \cdot \vec{x} = \|\vec{x}\|_2^2$  is  $2\vec{x}$ . Therefore the tangent space of the sphere  $S^{n-1}$  is orthogonal to  $\vec{x}$ , and the normal vector is  $\vec{n} = \frac{\vec{x}}{\|\vec{x}\|_2}$ .

**Example 8.4.3.** We try to characterize continuously differentiable functions  $f(x, y)$  satisfying  $xf_x = yf_y$ . Note that the condition is the same as  $\nabla f \perp (x, -y)$ , which is equivalent to  $\nabla f$  being parallel to  $(y, x) = \nabla(xy)$ . This means that the levels of  $f$  and the levels of  $xy$  are tangential everywhere. Therefore we expect the levels of  $f(x, y)$  and  $xy$  to be the same, and we should have  $f(x, y) = h(xy)$  for some continuously differentiable  $h(t)$ .

For the rigorous argument, let  $f(x, y) = g(x, xy)$ , or  $g(x, z) = f\left(x, \frac{z}{x}\right)$ , for the case  $x \neq 0$ . Then

$$g_x = f_x - \frac{z}{x^2}f_y = \frac{1}{x}\left(xf_x - \frac{z}{x}f_y\right) = 0.$$

This shows that  $g$  is independent of the first variable, and we have  $f(x, y) = h(xy)$ . For the case  $y \neq 0$ , we may get the same conclusion by defining  $f(x, y) = g(xy, y)$ .

A vast extension of the discussion is the theory of functional dependence. See Exercises 8.119 through 8.124.

**Exercise 8.59.** Extend Example 8.4.1 to hypersurface in  $\mathbb{R}^n$  given by the graph of a function on an open subset of  $\mathbb{R}^{n-1}$ .

**Exercise 8.60.** Find continuously differentiable functions satisfying the equations.

1.  $af_x = bf_y$ .
2.  $xf_x + yf_y = 0$ .
3.  $f_x = f_y = f_z$ .
4.  $xf_x = yf_y = zf_z$ .

**Exercise 8.61.** Suppose  $\vec{a}$  is a nonzero vector in  $\mathbb{R}^n$ . Suppose  $\vec{b}_1, \vec{b}_2, \dots, \vec{b}_{n-1}$  are linearly independent vectors orthogonal to  $\vec{a}$ . Prove that a function  $f$  on whole  $\mathbb{R}^n$  (or on any open convex subset) satisfies  $D_{\vec{a}}f = 0$  if and only if  $f(\vec{x}) = g(\vec{b}_1 \cdot \vec{x}, \vec{b}_2 \cdot \vec{x}, \dots, \vec{b}_{n-1} \cdot \vec{x})$  for some function  $g$  on  $\mathbb{R}^{n-1}$ .

## Submanifold of Euclidean Space

We have been vague about how “nice” a subset needs to be in order to become a hypersurface. At least we know that the graph of a differentiable function should be considered as a hypersurface. In fact, we wish to describe nice  $k$ -dimensional subsets in  $\mathbb{R}^n$ , where  $k$  is not necessarily  $n - 1$ . The graph of a map  $F: \mathbb{R}^k \rightarrow \mathbb{R}^{n-k}$

$$\{(\vec{x}, F(\vec{x})) : \vec{x} \in \mathbb{R}^k\} \subset \mathbb{R}^n$$

should be considered as a nice  $k$ -dimensional subset in  $\mathbb{R}^n$ . In general, however, we should not insist that it is always the last  $n - k$  coordinates that can be expressed as a map of the first  $k$  coordinates.

**Definition 8.4.1.** A  $k$ -dimensional (differentiable) submanifold of  $\mathbb{R}^n$  is a subset  $M$ , such that near any point on  $M$ , the subset  $M$  is the graph of a continuously differentiable map of some choice of  $n - k$  coordinates in terms of the other  $k$  coordinates.

**Example 8.4.4.** The unit circle  $S^1 = \{(x, y) : x^2 + y^2 = 1\} \subset \mathbb{R}^2$  is the graph of a function of  $y$  in  $x$  near  $(x_0, y_0) \in S^1$  if  $y_0 \neq 0$ . In fact, the function is given by  $y = \pm\sqrt{1 - x^2}$ ,

where the sign is the same as the sign of  $y_0$ . The unit circle is also the graph of a function of  $x$  in  $y$  near  $(x_0, y_0)$  if  $x_0 \neq 0$ . Therefore the unit circle is a 1-dimensional submanifold of  $\mathbb{R}^2$ .

**Example 8.4.5.** Suppose a parameterized curve  $\phi(t) = (x(t), y(t), z(t))$  is continuously differentiable, and the tangent vector  $\phi'(t)$  is never zero. At a point  $\phi(t_0)$  on the curve, at least one coordinate of  $\phi'(t_0)$  must be nonzero, say  $x'(t_0) \neq 0$ . Then by the continuity of  $x'$ , we have  $x' > 0$  near  $t_0$  or  $x' < 0$  near  $t_0$ . Therefore  $x(t)$  is strictly monotone and must be invertible near  $t_0$ . By Proposition 3.2.2,  $x = x(t)$  has differentiable inverse  $t = t(x)$  near  $x_0 = x(t_0)$ , and  $t'(x) = \frac{1}{x'(t(x))}$  is continuous. Therefore the curve may be reparameterized by  $x$  near  $\phi(t_0)$

$$\psi(x) = \phi(t(x)) = (x, y(t(x)), z(t(x))).$$

In other words, the coordinates  $y$  and  $z$  are continuously differentiable functions of the coordinate  $x$ . Similarly if  $y'(t_0) \neq 0$  or  $z'(t_0) \neq 0$ , then near  $\phi(t_0)$ ,  $x, z$  are continuous functions of  $y$ , or  $(x, y)$  are continuous functions of  $z$ . We conclude that the parameterized curve is a 1-dimensional submanifold.

Of course the argument is not restricted to curves in  $\mathbb{R}^3$ . If a continuously differentiable parameterized curve  $\phi(t)$  in  $\mathbb{R}^n$  is *regular*, in the sense that  $\phi'(t) \neq \vec{0}$  for any  $t$ , then the curve is a 1-dimensional submanifold of  $\mathbb{R}^n$ .

For a specific example, the parameterized unit circle  $\phi(t) = (\cos t, \sin t)$  has the tangent  $\phi'(t) = (-\sin t, \cos t)$ . If  $x' = -\sin t \neq 0$  (i.e., away from  $(1, 0)$  and  $(-1, 0)$ ), then  $x(t) = \cos t$  has local inverse  $t = \arccos x$ , so that  $y = \sin t = \sin(\arccos x) = \pm\sqrt{1-x^2}$  is a continuously differentiable function of  $x$ . Similarly, away from  $(0, 1)$  and  $(0, -1)$ ,  $x = \cos(\arcsin y) = \pm\sqrt{1-y^2}$  is a continuously differentiable function of  $y$ . Therefore the parameterized unit circle is a submanifold.

**Example 8.4.6.** A parameterized surface  $\sigma(u, v)$  is *regular* if the tangent vectors  $\sigma_u$  and  $\sigma_v$  are not parallel. In other words, they are linearly independent.

For the parameterized sphere (6.2.1), we have

$$\frac{\partial(x, y, z)}{\partial(\phi, \theta)} = (\sigma_\phi \quad \sigma_\theta) = \begin{pmatrix} \cos \phi \cos \theta & -\sin \phi \sin \theta \\ \cos \phi \sin \theta & \sin \phi \cos \theta \\ -\sin \phi & 0 \end{pmatrix}.$$

The parameterized sphere is regular if and only if the matrix has rank 2. This happens exactly when  $\sin \phi \neq 0$ , i.e., away from the north pole  $(0, 0, 1)$  and the south pole  $(0, 0, -1)$ .

If  $\sin \phi \neq 0$  and  $\cos \phi \neq 0$  (i.e., away from the two poles and the equator), then the first two rows  $\frac{\partial(x, y)}{\partial(\phi, \theta)}$  of  $\frac{\partial(x, y, z)}{\partial(\phi, \theta)}$  is invertible. By the Inverse Function Theorem (Theorem 8.3.1), the map  $(\phi, \theta) \mapsto (x, y)$  is locally invertible. In fact, the inverse is given by  $\phi = \arcsin \sqrt{x^2 + y^2}$ ,  $\theta = \arctan \frac{y}{x}$ . Substituting into  $z = \cos \phi$ , we get  $z = \sqrt{1-x^2-y^2}$  on the northern hemisphere and  $z = -\sqrt{1-x^2-y^2}$  on the southern hemisphere.

If  $\sin \phi \neq 0$  and  $\sin \theta \neq 0$ , then  $\frac{\partial(x, z)}{\partial(\phi, \theta)}$  is invertible, and the Inverse Function Theorem tells us that the map  $(\phi, \theta) \mapsto (x, z)$  is locally invertible. Substituting the local inverse  $\phi = \phi(x, z)$ ,  $\theta = \theta(x, z)$  into  $y = \sin \phi \sin \theta$ , we see that  $y$  can be locally written as a continuously differentiable function of  $x$  and  $z$ . Similarly, if  $\sin \phi \neq 0$  and  $\cos \theta \neq 0$ , then



$\frac{\partial(y, z)}{\partial(\phi, \theta)}$  is invertible, and  $x$  can be locally written as a continuously differentiable function of  $y$  and  $z$ .

The discussion covers all the points on the sphere except the two poles, and shows that the sphere is a submanifold except at the two poles. Moreover, near the two poles, the sphere is the graph of continuously differentiable functions  $\pm\sqrt{1-x^2-y^2}$ . Therefore the whole sphere is a submanifold.

**Exercise 8.62.** Find the places where the cycloid  $\phi(t) = (t - \sin t, 1 - \cos t)$  is regular and then compute the partial derivative of one coordinate with respect to the other coordinate.

**Exercise 8.63.** Find the places where the parameterized torus (6.2.2) is regular and then compute the partial derivatives of one coordinate with respect to the other two coordinates at regular points.

**Exercise 8.64.** Suppose  $\phi(t) = (x(t), y(t))$  is a regular parameterized curve in  $\mathbb{R}^2$ . By revolving the curve with respect to the  $x$ -axis, we get a surface of revolution

$$\sigma(t, \theta) = (x(t), y(t) \cos \theta, y(t) \sin \theta).$$

Is  $\sigma$  a regular parameterized surface?

The argument in Examples 8.4.5 and 8.4.6 is quite general, and can be summarized as follows.

**Proposition 8.4.2.** Suppose  $F: \mathbb{R}^k \rightarrow \mathbb{R}^n$  is a continuously differentiable map on an open subset  $U$  of  $\mathbb{R}^k$ . If the derivative  $F'(\vec{u}): \mathbb{R}^k \rightarrow \mathbb{R}^n$  is injective for any  $\vec{u} \in U$ , then the image

$$M = F(U) = \{F(\vec{u}) : \vec{u} \in U\}$$

is a  $k$ -dimensional submanifold of  $\mathbb{R}^n$ .

The map in the proposition is called a *regular parameterization* of the submanifold  $M$ . In general, to show that a subset  $M \subset \mathbb{R}^n$  is a submanifold, we may need to use several regular parameterizations, each covering a piece of  $M$ , and all the pieces together covering the whole  $M$ .

Let  $F = (f_1, f_2, \dots, f_n)$ . Then  $F(U)$  is given by

$$x_i = f_i(u_1, u_2, \dots, u_k), \quad (u_1, u_2, \dots, u_k) \in U.$$

The injectivity condition means that the  $n \times k$  Jacobian matrix

$$\frac{\partial(x_1, x_2, \dots, x_n)}{\partial(u_1, u_2, \dots, u_k)} = \begin{pmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \cdots & \frac{\partial x_1}{\partial u_k} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \cdots & \frac{\partial x_2}{\partial u_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \cdots & \frac{\partial x_n}{\partial u_k} \end{pmatrix}$$

has full rank  $k$ . Linear algebra tells us that some choice of  $k$  rows from the matrix form an invertible matrix. Then the Inverse Function Theorem implies that the  $k$  coordinates corresponding to these  $k$  rows can be considered as an invertible map of  $(u_1, u_2, \dots, u_k)$ . In other words,  $u_i$  can be locally expressed as continuously differentiable functions of these  $k$  coordinates. Then after substituting these expressions, the other  $n - k$  coordinates can be locally expressed as continuously differentiable functions of these  $k$  coordinates.

## Level Submanifold

The discussion of hypersurface suggests that a level of a function  $f(\vec{x})$  will be an  $(n - 1)$ -dimensional submanifold of  $\mathbb{R}^n$  as long as  $\nabla f$  is never zero. For example, consider the level  $S = \{(x, y, z): f(x, y, z) = 0\}$  of a continuously differentiable function. Suppose we have  $\nabla f(\vec{x}_0) \neq \vec{0}$  for some  $\vec{x}_0 = (x_0, y_0, z_0)$  in  $S$ . Then some coordinate of  $\nabla f(\vec{x}_0)$  is nonzero, say  $f_x(x_0, y_0, z_0) \neq 0$ . and the Implicit Function Theorem (Theorem 8.3.2) says that  $f(x, y, z) = 0$  defines  $x$  as a continuously differentiable function of  $(y, z)$  near  $\vec{x}_0$ . Therefore  $S$  is a manifold near  $\vec{x}_0$ . If  $\nabla f \neq \vec{0}$  on the whole  $S$ , then the argument shows that  $S$  is a 2-dimensional submanifold of  $\mathbb{R}^3$ .

In general, we consider several equalities

$$f_i(x_1, x_2, \dots, x_n) = c_i, \quad i = 1, 2, \dots, k,$$

which is the intersection of  $k$  levels of  $k$  functions. We expect that each equation cuts down the freedom by 1, and we should get an  $(n - k)$ -dimensional submanifold.

Denote  $F = (f_1, f_2, \dots, f_k): \mathbb{R}^n \rightarrow \mathbb{R}^k$  and  $\vec{c} = (c_1, c_2, \dots, c_k)$ . The argument for the level surface in  $\mathbb{R}^3$  can be extended.

**Proposition 8.4.3.** *Suppose  $F: \mathbb{R}^n \rightarrow \mathbb{R}^k$  is a continuously differentiable map on an open subset  $U$  of  $\mathbb{R}^n$ . If the derivative  $F'(\vec{x}): \mathbb{R}^n \rightarrow \mathbb{R}^k$  is surjective for any  $\vec{x} \in U$  satisfying  $F(\vec{x}) = \vec{c}$ , then the preimage*

$$M = F^{-1}(\vec{c}) = \{\vec{x} \in U: F(\vec{x}) = \vec{c}\}$$

*is an  $(n - k)$ -dimensional submanifold in  $\mathbb{R}^n$ .*

When the subjectivity condition is satisfied, we say  $\vec{c}$  is a *regular value* of the map  $F$ . The condition means that the gradients  $\nabla f_1, \nabla f_2, \dots, \nabla f_k$  are linearly independent at every point of  $M$ .

For a regular value  $\vec{c}$ , the  $n \times k$  Jacobian matrix

$$\frac{\partial(f_1, f_2, \dots, f_k)}{\partial(x_1, x_2, \dots, x_n)} = \begin{pmatrix} \nabla f_1 \\ \nabla f_2 \\ \vdots \\ \nabla f_k \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \frac{\partial f_k}{\partial x_2} & \cdots & \frac{\partial f_k}{\partial x_n} \end{pmatrix}$$

has full rank  $k$ . Linear algebra tells us that some choice of  $k$  columns from the matrix form an invertible matrix. By the Implicit Function Theorem (Theorem 8.3.2, applied to  $F(\vec{x}) - \vec{c} = \vec{0}$ ), the  $k$  variables corresponding to these columns can be locally expressed as continually differentiable functions of the other  $n - k$  variables.

**Example 8.4.7.** The unit sphere  $S^{n-1} = f^{-1}(1)$ , where  $f(\vec{x}) = \vec{x} \cdot \vec{x} = x_1^2 + x_2^2 + \cdots + x_n^2$ . The number 1 is a regular value because

$$f(\vec{x}) = 1 \implies \|\vec{x}\|_2 \neq 0 \implies \vec{x} \neq \vec{0} \implies \nabla f = 2\vec{x} \neq \vec{0}.$$

Therefore at any point of the sphere, we have some  $x_i \neq 0$ . If  $x_i > 0$ , then  $f(\vec{x}) = 1$  has continuously differentiable solution

$$x_i = \sqrt{1 - x_1^2 - \cdots - x_{i-1}^2 - x_{i+1}^2 - \cdots - x_n^2}.$$

If  $x_i < 0$ , then  $f(\vec{x}) = 1$  has continuously differentiable solution

$$x_i = -\sqrt{1 - x_1^2 - \cdots - x_{i-1}^2 - x_{i+1}^2 - \cdots - x_n^2}.$$

This shows that the sphere is a submanifold.

**Example 8.4.8.** The intersection of the unit sphere  $f(x, y, z) = x^2 + y^2 + z^2 = 1$  and the plane  $g(x, y, z) = x + y + z = a$  is supposed to be a circle and therefore should be a submanifold. The intersection can be described as  $F^{-1}(1, a)$ , where

$$F(x, y, z) = (f(x, y, z), g(x, y, z)) = (x^2 + y^2 + z^2, x + y + z): \mathbb{R}^3 \rightarrow \mathbb{R}^2.$$

We need to find out when  $(1, a)$  is a regular value of  $F$ . The irregularity happens when  $\nabla f = 2(x, y, z)$  and  $\nabla g = (1, 1, 1)$  are parallel, which means  $x = y = z$ . Combined with  $x^2 + y^2 + z^2 = 1$  and  $x + y + z = a$  (because the regularity is only verified for those points in  $F^{-1}(1, a)$ ), the irregularity means either

$$x = y = z = \frac{1}{\sqrt{3}}, \quad a = \sqrt{3},$$

or

$$x = y = z = -\frac{1}{\sqrt{3}}, \quad a = -\sqrt{3}.$$

Then we conclude that the intersection is a submanifold when  $a \neq \pm\sqrt{3}$ . In fact, the intersection is indeed a circle when  $|a| < \sqrt{3}$ , and the intersection is empty when  $|a| > \sqrt{3}$ .

**Example 8.4.9.** Consider the collection  $2 \times 2$  matrices of determinant 1

$$M = \left\{ X = \begin{pmatrix} x & y \\ z & w \end{pmatrix} : \det X = 1 \right\}.$$

This can be considered as a subset of  $\mathbb{R}^4$

$$M = \{(x, y, z, w) : xw - yz = 1\} = f^{-1}(1), \quad f(x, y, z, w) = xw - yz.$$

The number 1 is a regular value because

$$f(x, y, z, w) = 1 \implies (x, y, z, w) \neq \vec{0} \implies \nabla f = (w, -z, -y, x) \neq \vec{0}.$$

Therefore  $M$  is a 3-dimensional submanifold of  $\mathbb{R}^4$ .

**Exercise 8.65.** Find the condition such that the intersection of the ellipsoid  $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = \lambda$  and the hyperboloid  $\frac{x^2}{A^2} + \frac{y^2}{B^2} - \frac{z^2}{C^2} = \mu$  is a curve in  $\mathbb{R}^3$ .

**Exercise 8.66.** Let  $S(\vec{a}, r)$  be the sphere of  $\mathbb{R}^n$  centered at  $\vec{a}$  and has radius  $r$ . Find when the intersection of  $S(\vec{a}, r)$  and  $S(\vec{b}, s)$  is an  $(n-2)$ -dimensional submanifold. Moreover, when is the intersection of three spheres an  $(n-3)$ -dimensional submanifold?

**Exercise 8.67.** The *special linear group* is the collection of square matrices of determinant 1

$$SL(n) = \{n \times n \text{ matrix } X : \det X = 1\}.$$

Example 8.4.9 shows that  $SL(2)$  is a submanifold of  $\mathbb{R}^4$ . Prove that  $SL(n)$  is an  $(n^2 - 1)$ -dimensional submanifold of  $\mathbb{R}^{n^2}$ . What if 1 is changed to another number?

**Exercise 8.68.** The *orthogonal group* is the collection of orthogonal matrices

$$O(n) = \{n \times n \text{ matrix } X : X^T X = I\}.$$

By considering the map  $F(X) = X^T X : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{\frac{n(n+1)}{2}}$  in Exercise 8.10, where  $\mathbb{R}^{\frac{n(n+1)}{2}}$  is the space of  $n \times n$  symmetric matrices, prove that  $O(n)$  is an  $\frac{n(n-1)}{2}$ -dimensional submanifold of  $\mathbb{R}^{n^2}$ .

## Tangent Space

The *tangent space* of a submanifold  $M \subset \mathbb{R}^n$  is the collection of the tangent vectors of all curves in  $M$

$$T_{\vec{x}_0} M = \{\phi'(0) : \phi(t) \in M \text{ for all } t, \phi(0) = \vec{x}_0\}.$$

If  $M = F(U)$  is given by a regular parameterization in Proposition 8.4.2, then

$$\begin{aligned} T_{F(\vec{u}_0)} F(U) &= \{(F \circ \phi)'(0) : \phi(t) \text{ is a curve in } U, \phi(0) = \vec{u}_0\} \\ &= \{F'(\vec{u}_0)(\phi'(0)) : \phi(t) \text{ is a curve in } U, \phi(0) = \vec{u}_0\} \\ &= \{F'(\vec{u}_0)(\vec{v}) : \vec{v} \in \mathbb{R}^k\} \end{aligned} \quad (8.4.1)$$

is the image of the derivative  $F'(\vec{u}_0) : \mathbb{R}^k \rightarrow \mathbb{R}^n$ . If  $M = F^{-1}(\vec{c})$  is the preimage of a regular value in Proposition 8.4.3, then

$$\begin{aligned} T_{\vec{x}_0} F^{-1}(\vec{c}) &= \{\phi'(0) : F(\phi(t)) = \vec{c}, \phi(0) = \vec{x}_0\} \\ &= \{\phi'(0) : F(\phi(t))' = F'(\phi(t))(\phi'(t)) = \vec{0}, \phi(0) = \vec{x}_0\} \\ &= \{\phi'(0) : F'(\vec{x}_0)(\phi'(0)) = \vec{0}\} \\ &= \{\vec{v} : F'(\vec{x}_0)(\vec{v}) = \vec{0}\} \end{aligned} \quad (8.4.2)$$

is the kernel of the derivative  $F'(\vec{x}_0) : \mathbb{R}^n \rightarrow \mathbb{R}^k$ . If  $F = (f_1, f_2, \dots, f_k)$ , then this means

$$T_{\vec{x}_0} F^{-1}(\vec{c}) = \{\vec{v} : \nabla f_1 \cdot \vec{v} = \nabla f_2 \cdot \vec{v} = \dots = \nabla f_k \cdot \vec{v} = 0\}.$$

Therefore the gradients span the *normal space* of the submanifold.

**Example 8.4.10.** From Example 8.4.7, the tangent space of the unit sphere is the vectors orthogonal to the gradient  $2\vec{x}$

$$T_{\vec{x}}S^{n-1} = \{\vec{v}: \vec{v} \cdot \vec{x} = 0\}.$$

The tangent space of the circle in Example 8.4.8 consists of vectors  $\vec{v} = (u, v, w)$  satisfying

$$\nabla f \cdot \vec{v} = 2(xu + yv + zw) = 0, \quad \nabla g \cdot \vec{v} = u + v + w = 0.$$

This means that  $\vec{v}$  is parallel to  $(y - z, z - x, x - y)$ . So the tangent space is the straight line spanned by  $(y - z, z - x, x - y)$ .

**Exercise 8.69.** Find the the tangent space of the torus (6.2.2).

**Exercise 8.70.** Find the regular values of the map and determine the tangent space of the preimage of regular value.

1.  $f(x, y) = x^2 + y^2 + z^2 + xy + yz + zx + x + y + z$ .
2.  $F(x, y, z) = (x + y + z, xy + yz + zx)$ .

**Exercise 8.71.** Find a regular value  $\lambda$  for  $xyz$ , so that the surface  $xyz = \lambda$  is tangential to the ellipse  $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$  at some point.

**Exercise 8.72.** Prove that any sphere  $x^2 + y^2 + z^2 = a^2$  and any cone  $x^2 + y^2 = b^2 z^2$  are orthogonal at their intersections. Can you extend this to higher dimension?

## 8.5 High Order Approximation

A map  $P: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a *polynomial map* of degree  $k$  if all its coordinate functions are polynomials of degree  $k$ . A map  $F(\vec{x})$  defined near  $\vec{x}_0$  is  *$k$ -th order differentiable* at  $\vec{x}_0$  if it is approximated by a polynomial map  $P$  of degree  $k$ . In other words, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|\Delta\vec{x}\| < \delta \implies \|F(\vec{x}) - P(\vec{x})\| \leq \epsilon \|\Delta\vec{x}\|^k.$$

The approximate polynomial can be written as

$$P(\vec{x}) = F(\vec{x}_0) + F'(\vec{x}_0)(\Delta\vec{x}) + \frac{1}{2}F''(\vec{x}_0)(\Delta\vec{x}) + \cdots + \frac{1}{k!}F^{(k)}(\vec{x}_0)(\Delta\vec{x}),$$

where the coordinates of  $F^{(i)}(\vec{x}_0)$  are  $i$ -th order forms. Then  $F^{(i)}(\vec{x}_0)$  is the  *$i$ -th order derivative* of  $F$  at  $\vec{x}_0$ , and  $d^i F = F^{(i)}(\vec{x}_0)(d\vec{x})$  is the  *$i$ -th order differential*.

A map is  $k$ -th order differentiable if and only if each coordinate is  $k$ -th order differentiable. A function  $f(\vec{x})$  is  $k$ -th order differentiable at  $\vec{x}_0$  if there is

$$p(\vec{x}) = \sum_{k_1 + \cdots + k_n \leq k, k_i \geq 0} b^{k_1 \cdots k_n} \Delta x_1^{k_1} \cdots \Delta x_n^{k_n},$$

such that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|\Delta\vec{x}\| < \delta \implies |f(\vec{x}) - p(\vec{x})| \leq \epsilon \|\Delta\vec{x}\|^k.$$

The  $k$ -th order derivative

$$f^{(k)}(\vec{x}_0)(\vec{v}) = k! \sum_{k_1 + \dots + k_n = k, k_i \geq 0} b^{k_1 \dots k_n} v_1^{k_1} \dots v_n^{k_n}$$

is a  $k$ -th order form, and the  $k$ -th order differential is

$$d^k f = k! \sum_{k_1 + \dots + k_n = k, k_i \geq 0} b^{k_1 \dots k_n} dx_1^{k_1} \dots dx_n^{k_n}.$$

**Exercise 8.73.** Extend Exercise 8.33 to high order. Suppose  $F$  is  $k$ -th order differentiable at  $\vec{x}_0$ , and  $F^{(i)}(\vec{x}_0)$  vanishes for  $0 \leq i \leq k-1$ . Suppose  $G$  is  $l$ -th order differentiable at  $\vec{x}_0$ , and  $F^{(j)}(\vec{x}_0)$  vanishes for  $0 \leq j \leq l-1$ .

1. Prove that for any bilinear map  $B$ , we have

$$B(F, G) \sim_{\|\Delta \vec{x}\|^{k+l}} \frac{1}{k!l!} B(F^{(k)}(\vec{x}_0)(\Delta \vec{x}), G^{(l)}(\vec{x}_0)(\Delta \vec{x})).$$

2. Prove that for any quadratic map  $Q$ , we have

$$Q(F) \sim_{\|\Delta \vec{x}\|^{2k}} Q(F^{(k)}(\vec{x}_0)(\Delta \vec{x})).$$

Moreover, further extend to multilinear functions of more variables.

## Second Order Differentiation

A function  $f(\vec{x})$  defined near  $\vec{x}_0 \in \mathbb{R}^n$  is *second order differentiable* at  $\vec{x}_0$  if it is approximated by a quadratic function

$$\begin{aligned} p(\vec{x}) &= a + \sum_{1 \leq i \leq n} b_i(x_i - x_{i0}) + \sum_{1 \leq i, j \leq n} c_{ij}(x_i - x_{i0})(x_j - x_{j0}) \\ &= a + \vec{b} \cdot \Delta \vec{x} + C \Delta \vec{x} \cdot \Delta \vec{x}. \end{aligned}$$

This means that, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|\Delta \vec{x}\| < \delta \implies |f(\vec{x}) - p(\vec{x})| \leq \epsilon \|\Delta \vec{x}\|^2.$$

The *second order derivative* is the quadratic form

$$f''(\vec{x}_0)(\vec{v}) = 2 \sum_{1 \leq i, j \leq n} c_{ij} v_i v_j = 2C\vec{v} \cdot \vec{v},$$

and the *second order differential* is

$$d^2 f = 2 \sum_{1 \leq i, j \leq n} c_{ij} dx_i dx_j = 2Cd\vec{x} \cdot d\vec{x}.$$

**Example 8.5.1.** By  $\|\vec{x}\|_2^2 = \|\vec{x}_0\|_2^2 + 2\vec{x}_0 \cdot \Delta \vec{x} + \|\Delta \vec{x}\|_2^2$ , we have the first order derivative  $(\|\vec{x}\|_2^2)'(\vec{x}_0)(\vec{v}) = 2\vec{x}_0 \cdot \vec{v}$  and the second order derivative  $(\|\vec{x}\|_2^2)''(\vec{x}_0)(\vec{v}) = 2\|\vec{v}\|_2^2$ .

In general, the second order derivative of a quadratic form  $q(\vec{x})$  is  $q''(\vec{x}_0)(\vec{v}) = 2q(\vec{v})$ .

**Example 8.5.2.** The computation in Example 8.1.5 tells us that the second order derivative of the matrix square map  $F(X) = X^2: \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$  is  $F''(A)(H) = 2H^2$ .

**Example 8.5.3.** The computation in Example 8.1.6 also gives the quadratic approximation of  $\det$  near  $I$ . The second order terms in the quadratic approximations are

$$\det(\vec{e}_1 \cdots \vec{h}_i \cdots \vec{h}_j \cdots \vec{e}_n) = \det \begin{pmatrix} h_{ii} & h_{ij} \\ h_{ji} & h_{jj} \end{pmatrix} = h_{ii}h_{jj} - h_{ij}h_{ji}.$$

Therefore the second order derivative

$$\det''(I)(H) = 2 \sum_{1 \leq i < j \leq n} \det \begin{pmatrix} h_{ii} & h_{ij} \\ h_{ji} & h_{jj} \end{pmatrix}.$$

**Exercise 8.74.** Find the second order derivative of a bilinear map  $B: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^k$ . What about multilinear maps?

**Exercise 8.75.** Find the second order derivative of the  $k$ -th power map of matrices.

**Exercise 8.76.** Find the second order derivative of the inverse matrix map at  $I$ .

## Second Order Partial Derivative

Similar to the single variable case, the second order differentiability implies the first order differentiability, so that  $a = f(\vec{x}_0)$  and  $b_i = \frac{\partial f(\vec{x}_0)}{\partial x_i}$ . We expect the coefficients  $c_{ij}$  to be the second order partial derivatives

$$c_{ij} = \frac{1}{2} f_{x_j x_i}(\vec{x}_0) = \frac{1}{2} \frac{\partial^2 f(\vec{x}_0)}{\partial x_i \partial x_j} = \frac{1}{2} \frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_j} \right)_{\vec{x}=\vec{x}_0}.$$

Example 3.4.8 suggests that the second order differentiability alone should not imply the existence of the second order partial derivatives. For example, for  $p > 2$ , the function  $\|\vec{x}\|_2^p D(\|\vec{x}\|_2)$  is second order differentiable at  $\vec{0}$ , with the zero function as the quadratic approximation. However, the function has no first order partial derivatives away from the origin, so that the second order partial derivatives are not defined at the origin.

Conversely, Example 8.1.8 shows that the existence of first order partial derivatives does not imply the differentiability. So we do not expect the existence of second order partial derivatives alone to imply the second order differentiability. See Example 8.5.4. On the other hand, the multivariable version of Theorem 3.4.2 is true under slightly stronger condition. The following is the quadratic case of the multivariable version of Theorem 3.4.2. The general case will be given by Theorem 8.5.3.

**Theorem 8.5.1.** Suppose a function  $f$  is differentiable near  $\vec{x}_0$  and the partial derivatives  $\frac{\partial f}{\partial x_i}$  are differentiable at  $\vec{x}_0$ . Then  $f$  is second order differentiable at

$\vec{x}_0$ , with the quadratic approximation

$$T_2(\vec{x}) = f(\vec{x}_0) + \sum_{1 \leq i \leq n} \frac{\partial f(\vec{x}_0)}{\partial x_i} \Delta x_i + \frac{1}{2} \sum_{1 \leq i, j \leq n} \frac{\partial^2 f(\vec{x}_0)}{\partial x_i \partial x_j} \Delta x_i \Delta x_j.$$

*Proof.* We only prove the case  $n = 2$  and  $\vec{x}_0 = (0, 0)$ . The general case is similar. To show that

$$\begin{aligned} p(x, y) &= f(0, 0) + f_x(0, 0)x + f_y(0, 0)y \\ &\quad + \frac{1}{2}(f_{xx}(0, 0)x^2 + f_{xy}(0, 0)xy + f_{yx}(0, 0)yx + f_{yy}(0, 0)y^2) \end{aligned}$$

approximates  $f$  near  $(0, 0)$ , we restrict the remainder  $R_2(x, y) = f(x, y) - p(x, y)$  to the straight lines passing through  $(0, 0)$ . Therefore for fixed  $(x, y)$  close to  $(0, 0)$ , we introduce

$$r_2(t) = R_2(tx, ty) = f(tx, ty) - p(tx, ty), \quad t \in [0, 1].$$

Inspired by the proof of Theorem 3.4.2, we apply Cauchy's Mean Value Theorem to get

$$R_2(x, y) = r_2(1) = \frac{r_2(1) - r_2(0)}{1^2 - 0^2} = \frac{r_2'(c)}{2c}, \quad 0 < c < 1.$$

To compute  $r_2'(c)$ , we note that the differentiability of  $f$  near  $(0, 0)$  allows us to use the chain rule to get

$$\begin{aligned} r_2'(t) &= f_x(tx, ty)x + f_y(tx, ty)y - f_x(0, 0)x - f_y(0, 0)y \\ &\quad - (f_{xx}(0, 0)x^2 + f_{xy}(0, 0)xy + f_{yx}(0, 0)yx + f_{yy}(0, 0)y^2)t \\ &= x[f_x(tx, ty) - f_x(0, 0) - f_{xx}(0, 0)tx - f_{xy}(0, 0)ty] \\ &\quad + y[f_y(tx, ty) - f_y(0, 0) - f_{yx}(0, 0)tx - f_{yy}(0, 0)ty]. \end{aligned}$$

Further by the differentiability of  $f_x$  and  $f_y$ , for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|(\xi, \eta)\| < \delta$  implies

$$\begin{aligned} |f_x(\xi, \eta) - f_x(0, 0) - f_{xx}(0, 0)\xi - f_{xy}(0, 0)\eta| &< \epsilon\|(\xi, \eta)\|, \\ |f_y(\xi, \eta) - f_y(0, 0) - f_{yx}(0, 0)\xi - f_{yy}(0, 0)\eta| &< \epsilon\|(\xi, \eta)\|. \end{aligned}$$

Taking  $(\xi, \eta) = (cx, cy)$ , by the formula above for  $r_2'(t)$ , we find  $\|(x, y)\| < \delta$  implies

$$|R_2(x, y)| \leq \frac{|x|\epsilon\|(cx, cy)\| + |y|\epsilon\|(cx, cy)\|}{2c} \leq \frac{1}{2}\epsilon(|x| + |y|)\|(x, y)\|.$$

By the equivalence of norms, the right side is comparable to  $\epsilon\|(x, y)\|^2$ . Therefore we conclude that  $p(x, y)$  is a quadratic approximation of  $f(x, y)$  near  $(0, 0)$ .  $\square$



**Example 8.5.4.** Consider

$$f(x, y) = \begin{cases} \frac{x^2 y^2}{(x^2 + y^2)^2}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

We have

$$f_x = \begin{cases} \frac{2xy^2(x^2 - y^2)}{(x^2 + y^2)^3}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0), \end{cases} \quad f_y = \begin{cases} \frac{2x^2y(y^2 - x^2)}{(x^2 + y^2)^3}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

Then we further have  $f_{xx}(0, 0) = f_{yy}(0, 0) = f_{xy}(0, 0) = f_{yx}(0, 0) = 0$ . However, by Example 6.2.2, the function is not continuous at  $(0, 0)$ , let alone second order differentiable.

**Exercise 8.77.** Derive the chain rule for the second order derivative by composing the quadratic approximations.

**Exercise 8.78.** Prove that if  $f_{xy} = 0$  on a convex open subset, then  $f(x, y) = g(x) + h(y)$ . Can you extend this to more variables?

**Exercise 8.79.** Prove that if  $ff_{xy} = f_x f_y$  on a convex open subset, then  $f(x, y) = g(x)h(y)$ .

**Exercise 8.80.** Discuss the second order differentiability of a function  $f(\|\vec{x}\|_2)$  of the Euclidean norm.

**Exercise 8.81.** Discuss the existence of second order derivative and the second order differentiability of functions in Exercise 8.14.

## Order of Taking Partial Derivatives

The notation  $f_{xy}$  means first taking the partial derivative in  $x$  and then taking the partial derivative in  $y$ . In general, we do not expect  $f_{xy}$  and  $f_{yx}$  to be equal (see Example 8.5.5). On the other hand, under the assumption of Theorem 7.1.2, the second order partial derivatives are independent of the order of the variables.

**Theorem 8.5.2.** Suppose  $f(x, y)$  has partial derivatives  $f_x$  and  $f_y$  near  $(x_0, y_0)$ . If  $f_x$  and  $f_y$  are differentiable at  $(x_0, y_0)$ , then  $f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$ .

*Proof.* To simplify the notation, we may assume  $x_0 = 0 = y_0$ , without loss of generality. We note that  $\Delta x = x$  and  $\Delta y = y$ .

For fixed  $y$ , we apply the Mean Value Theorem to  $g(t) = f(t, y) - f(t, 0)$  for  $t$  between 0 and  $x$ . By the existence of  $f_x$  near  $(0, 0)$ , there is  $c$  between 0 and  $x$ , such that

$$\begin{aligned} f(x, y) - f(x, 0) - f(0, y) + f(0, 0) &= g(x) - g(0) = g'(c)x \\ &= (f_x(c, y) - f_x(c, 0))x. \end{aligned}$$

Next we use the differentiability of  $f_x$  at  $(0, 0)$  to estimate  $f_x(c, y)$  and  $f_x(c, 0)$ . For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|(x, y)\|_\infty < \delta$  implies

$$|f_x(x, y) - f_x(0, 0) - f_{xx}(0, 0)x - f_{xy}(0, 0)y| \leq \epsilon \|(x, y)\|_\infty.$$

Since  $\|(c, y) - (0, 0)\|_\infty \leq \|(x, y)\|_\infty$  and  $\|(c, 0) - (0, 0)\|_\infty \leq \|(x, y)\|_\infty$ , we may take  $(x, y)$  to be  $(c, y)$  or  $(c, 0)$  in the inequality above. Therefore  $\|(x, y)\|_\infty < \delta$  implies

$$\begin{aligned} |f_x(c, y) - f_x(0, 0) - f_{xx}(0, 0)c - f_{xy}(0, 0)y| &\leq \epsilon \|(x, y)\|_\infty, \\ |f_x(c, 0) - f_x(0, 0) - f_{xx}(0, 0)c| &\leq \epsilon \|(x, y)\|_\infty. \end{aligned}$$

Combining the two, we get

$$|f_x(c, y) - f_x(c, 0) - f_{xy}(0, 0)y| \leq 2\epsilon \|(x, y)\|_\infty,$$

and further get

$$\begin{aligned} &|f(x, y) - f(x, 0) - f(0, y) + f(0, 0) - f_{xy}(0, 0)xy| \\ &= |(f_x(c, y) - f_x(c, 0))x - f_{xy}(0, 0)xy| \\ &\leq 2\epsilon \|(x, y)\|_\infty |x| \leq 2\epsilon \|(x, y)\|_\infty^2. \end{aligned}$$

Exchanging  $x$  and  $y$ , we may use the existence of  $f_y$  near  $(0, 0)$  and the differentiability of  $f_y$  at  $(0, 0)$  to conclude that, for any  $\epsilon > 0$ , there is  $\delta' > 0$ , such that  $\|(x, y)\|_\infty < \delta'$  implies

$$|f(x, y) - f(0, y) - f(x, 0) + f(0, 0) - f_{yx}(0, 0)xy| \leq 2\epsilon \|(x, y)\|_\infty^2.$$

Combining the two estimations, we find that  $\|(x, y)\|_\infty < \min\{\delta, \delta'\}$  implies

$$|f_{xy}(0, 0)xy - f_{yx}(0, 0)xy| \leq 4\epsilon \|(x, y)\|_\infty^2.$$

By taking  $x = y$ , we find that  $|x| < \min\{\delta, \delta'\}$  implies

$$|f_{xy}(0, 0) - f_{yx}(0, 0)| \leq 4\epsilon.$$

Since  $\epsilon$  is arbitrary, we conclude that  $f_{xy}(0, 0) = f_{yx}(0, 0)$ . □

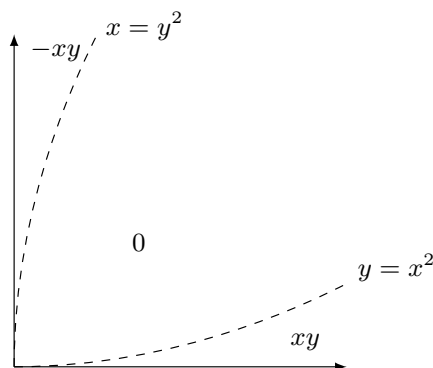
**Example 8.5.5.** We construct an example with  $f_{xy} \neq f_{yx}$ . Define (see Figure 8.5.1)

$$f(x, y) = \begin{cases} xy, & \text{if } |y| \leq x^2, \\ -xy, & \text{if } |x| \leq y^2, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that  $f_{xy}(0, 0)$  depends only on the value of  $f$  on the region  $|x| \leq y^2$ . Therefore  $f_{xy}(0, 0) = \frac{\partial^2(-xy)}{\partial y \partial x} = -1$ . Similarly, we have  $f_{yx}(0, 0) = \frac{\partial^2(xy)}{\partial x \partial y} = 1$ .

On the other hand, the function is second order differentiable, with 0 as the quadratic approximation. The reason is that for  $|y| \leq x^2 < 1$ , we have  $\|(x, y)\|_\infty = |x|$ , so that  $|f(x, y)| = |xy| \leq |x|^3 = \|(x, y)\|_\infty^3$ , and we also have  $|f(x, y)| \leq \|(x, y)\|_\infty^3$  on  $|x| \leq y^2$  for the similar reason. In fact, the formula in Theorem 7.1.2 for the quadratic approximation remains true.

Exercise 8.83 gives a function satisfying  $f_{xy}(0, 0) \neq f_{yx}(0, 0)$ , and is not second order differentiable.



**Figure 8.5.1.** an example with  $f_{xy} \neq f_{yx}$

**Exercise 8.82 (Clairaut).** Suppose  $f(x, y)$  has partial derivatives  $f_x, f_y, f_{xy}$  near  $(x_0, y_0)$  and  $f_{xy}$  is continuous at  $(x_0, y_0)$ .

1. Prove that the whole limit  $\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{f(x, y) - f(x, y_0) - f(x_0, y) + f(x_0, y_0)}{(x - x_0)(y - y_0)}$  converges.
2. Use Proposition 6.2.3 to prove that  $f_{yx}(x_0, y_0)$  exists and  $f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$ .

The assumption here is slightly different from Theorem 8.5.2, and is what you usually see in calculus textbooks. The proof here motivates the proof of Theorem 8.5.2.

**Exercise 8.83.** Consider the function

$$f(x, y) = \begin{cases} \frac{xy(x^2 - y^2)}{x^2 + y^2}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

1. Show that  $f$  has all the second order partial derivatives near  $(0, 0)$  but  $f_{xy}(0, 0) \neq f_{yx}(0, 0)$ .
2. By restricting respectively to  $x = 0, y = 0$  and  $x = y$ , show that if  $f(x, y)$  is second order differentiable, then the quadratic approximation must be 0.
3. Show that  $f$  is not second order differentiable at  $(0, 0)$ .

**Exercise 8.84.** By slightly modifying Example 8.5.5, construct a function that is second order differentiable at  $(0, 0)$ , has all second order partial derivatives, yet the Taylor expansion (the formula in Theorem 7.1.2) is not the quadratic approximation.

**Exercise 8.85.** Consider a second order polynomial

$$p(x_1, x_2) = a + b_1x_1 + b_2x_2 + c_{11}x_1^2 + c_{22}x_2^2 + 2c_{12}x_1x_2,$$

and numbers  $D_1, D_2, D_{11}, D_{22}, D_{12}, D_{21}$ . Prove the following are equivalent.

1. There is a function  $f(x_1, x_2)$  that is second order differentiable at  $(0, 0)$ , such that  $p(x, y)$  is the quadratic approximation of  $f(x, y)$  and  $f_{x_i}(0, 0) = D_i, f_{x_i x_j}(0, 0) = D_{ji}$ .
2.  $D_1 = b_1, D_2 = b_2, D_{11} = 2c_{11}, D_{22} = 2c_{22}$ .

Moreover, extend the result to more than two variables.

## Taylor Expansion

The discussion about quadratic approximations of multivariable functions easily extends to high order approximations. The  $k$ -th order partial derivative

$$\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} = D_{x_{i_1} x_{i_2} \cdots x_{i_k}} f = f_{x_{i_k} \cdots x_{i_2} x_{i_1}}$$

is obtained by successfully taking partial derivatives in  $x_{i_k}, \dots, x_{i_2}, x_{i_1}$ . Given all the partial derivatives up to order  $k$ , we may construct the  $k$ -th order Taylor expansion

$$T_k(\vec{x}) = \sum_{j=0}^k \frac{1}{j!} \sum_{1 \leq i_1, \dots, i_j \leq n} \frac{\partial^j f(\vec{x}_0)}{\partial x_{i_1} \cdots \partial x_{i_j}} \Delta x_{i_1} \cdots \Delta x_{i_j}.$$

The following is the high order version of Theorem 8.5.1 and the multivariable version of Theorem 3.4.2.

**Theorem 8.5.3.** *Suppose  $f(\vec{x})$  has continuous partial derivatives up to  $(k-1)$ -st order near  $\vec{x}_0$  and the  $(k-1)$ -st order partial derivatives are differentiable at  $\vec{x}_0$ . Then the  $k$ -th order Taylor expansion is the  $k$ -th order polynomial approximation at  $\vec{x}_0$ . In particular, the function is  $k$ -th order differentiable.*

By Theorem 8.5.2, the assumption already implies that the partial derivatives up to the  $k$ -th order are independent of the order of the variables. Therefore we can write

$$\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} = \frac{\partial^k f}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}},$$

where  $k_j$  is the number of  $x_j$  in the collection  $\{x_{i_1}, x_{i_2}, \dots, x_{i_k}\}$ , and the Taylor expansion becomes

$$T_k(\vec{x}) = \sum_{k_1 + \cdots + k_n \leq k, k_i \geq 0} \frac{1}{k_1! \cdots k_n!} \frac{\partial^{k_1 + \cdots + k_n} f(\vec{x}_0)}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}} \Delta x_1^{k_1} \cdots \Delta x_n^{k_n}.$$

A multivariable map is *continuously  $k$ -th order differentiable* if all the partial derivatives up to  $k$ -th order exist and are continuous. By Theorem 8.1.3, the conditions of Theorem 8.5.3 are satisfied, and the map  $\vec{x} \mapsto F^{(i)}(\vec{x})$  is continuous for all  $1 \leq i \leq k$ .

*Proof.* Restricting the remainder

$$R_k(\vec{x}) = f(\vec{x}) - T_k(\vec{x})$$

to the straight line connecting  $\vec{x}_0$  to  $\vec{x}$ , we get

$$r_k(t) = R_k((1-t)\vec{x}_0 + t\vec{x}) = R_k(\vec{x}_0 + t\Delta\vec{x}).$$

Since  $f(\vec{x})$  has continuous partial derivatives up to  $(k-1)$ -st order near  $\vec{x}_0$ , the partial derivatives of  $f(\vec{x})$  up to  $(k-2)$ -nd order are differentiable near  $\vec{x}_0$ . Then by

the chain rule,  $r_k$  has derivatives up to  $(k-1)$ -st order for  $t \in (-1, 1)$ . Moreover, we have

$$r_k(0) = r'_k(0) = r''_k(0) = \cdots = r_k^{(k-1)}(0) = 0.$$

By Cauchy's Mean Value Theorem, there are  $1 > c_1 > c_2 > \cdots > c_{k-1} > 0$ , such that

$$\begin{aligned} R_k(\vec{x}) &= r_k(1) = \frac{r_k(1) - r_k(0)}{1^k - 0^k} = \frac{r'_k(c_1)}{k c_1^{k-1}} = \frac{r'_k(c_1) - r'_k(0)}{k(c_1^{k-1} - 0^{k-1})} \\ &= \frac{r''_k(c_2)}{k(k-1)c_2^{k-2}} = \cdots = \frac{r_k^{(k-1)}(c_{k-1})}{k(k-1)\cdots 2c_{k-1}} = \frac{r_k^{(k-1)}(c_{k-1})}{k!c_{k-1}}. \end{aligned}$$

By the high order differentiability assumption, we may use the chain rule to get

$$r_k^{(k-1)}(t) = \sum_{1 \leq i_1, \dots, i_{k-1} \leq n} [\delta_{i_1, \dots, i_{k-1}}(t\Delta\vec{x}) - \lambda_{i_1, \dots, i_{k-1}}(t\Delta\vec{x})] \Delta x_{i_1} \cdots \Delta x_{i_{k-1}},$$

where

$$\delta_{i_1, \dots, i_{k-1}}(\Delta\vec{x}) = \frac{\partial^{k-1} f(\vec{x}_0 + \Delta\vec{x})}{\partial x_{i_1} \cdots \partial x_{i_{k-1}}}$$

is the  $(k-1)$ -st order partial derivative, and

$$\lambda_{i_1, \dots, i_{k-1}}(\Delta\vec{x}) = \frac{\partial^{k-1} f(\vec{x}_0)}{\partial x_{i_1} \cdots \partial x_{i_{k-1}}} + \sum_{1 \leq i \leq n} \frac{\partial^k f(\vec{x}_0)}{\partial x_i \partial x_{i_1} \cdots \partial x_{i_{k-1}}} \Delta x_i$$

is the linear approximation of  $\delta_{i_1, \dots, i_{k-1}}(\Delta\vec{x})$ . Since the  $(k-1)$ -st order partial derivatives  $\delta_{i_1, \dots, i_m}$  of  $f$  are differentiable, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|\Delta\vec{x}\| < \delta$  implies

$$|\delta_{i_1, \dots, i_{k-1}}(\Delta\vec{x}) - \lambda_{i_1, \dots, i_{k-1}}(\Delta\vec{x})| \leq \epsilon \|\Delta\vec{x}\|.$$

Then for  $\|\Delta\vec{x}\| < \delta$  and  $|t| < 1$ , we have

$$|r_k^{(k-1)}(t)| \leq \sum_{1 \leq i_1, \dots, i_{k-1} \leq n} \epsilon \|t\Delta\vec{x}\| |\Delta x_{i_1} \cdots \Delta x_{i_{k-1}}| \leq \epsilon n^{k-1} |t| \|\Delta\vec{x}\| \|\Delta\vec{x}\|_\infty^{k-1},$$

and

$$|R_k(\vec{x})| = \frac{|r_k^{(k-1)}(c_{k-1})|}{k!|c_{k-1}|} \leq \epsilon \frac{n^{k-1}}{k!} \|\Delta\vec{x}\| \|\Delta\vec{x}\|_\infty^{k-1}.$$

By the equivalence of norms, this implies  $\lim_{\Delta\vec{x} \rightarrow \vec{0}} \frac{R_k(\vec{x})}{\|\Delta\vec{x}\|^k} = 0$ . □

**Exercise 8.86.** Find high order differentiation.

1.  $x^y$  at  $(1, 4)$ , to third order.
2.  $\sin(x^2 + y^2)$  at  $(0, 0)$ , to fourth order.
3.  $x^y y^z$  at  $(1, 1, 1)$ , to third order.

4.  $\int_0^1 (1+x)^{t^2 y} dt$  at  $(0,0)$ , to third order.

**Exercise 8.87.** For a function  $f$  on  $\mathbb{R}^n$  and a linear transform  $L: \mathbb{R}^m \rightarrow \mathbb{R}^n$ . How are the high order derivatives of  $f$  and  $f \circ L$  related?

**Exercise 8.88.** Suppose  $y = y(\vec{x})$  is given implicitly by  $f(\vec{x}, y) = 0$ . Compute the third order derivatives of  $y$ .

**Exercise 8.89.** Find the high order derivatives of the determinant map at  $I$ .

**Exercise 8.90.** Find the third order derivative of the  $k$ -th power map of matrices.

**Exercise 8.91.** Find the high order derivatives of the inverse matrix map at  $I$ .

**Exercise 8.92.** Consider

$$f(x, y) = \begin{cases} \frac{x^k y^k}{(x^2 + y^2)^k}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

Show that  $f$  has all the partial derivatives up to  $k$ -th order. But  $f$  is not  $k$ -th order differentiable.

**Exercise 8.93.** Suppose  $f(\vec{x})$  has continuous partial derivatives up to  $(k+1)$ -st order near  $\vec{x}_0$ . Prove that for any  $\vec{x}$  near  $\vec{x}_0$ , there is  $\vec{c}$  on the straight line connecting  $\vec{x}_0$  to  $\vec{x}$ , such that

$$|f(\vec{x}) - T_k(\vec{x})| \leq \sum_{k_1 + \dots + k_n = k+1} \frac{1}{k_1! \dots k_n!} \left| \frac{\partial^{k+1} f(\vec{c})}{\partial x_1^{k_1} \dots \partial x_n^{k_n}} \right| |\Delta x_1|^{k_1} \dots |\Delta x_n|^{k_n}.$$

This is the multivariable version of Proposition 3.4.3.

## 8.6 Maximum and Minimum

A function  $f(\vec{x})$  has a *local maximum* at  $\vec{x}_0$  if there is  $\delta > 0$ , such that

$$\|\vec{x} - \vec{x}_0\| < \delta \implies f(\vec{x}) \leq f(\vec{x}_0).$$

In other words, the value of  $f$  at  $\vec{x}_0$  is biggest among the values of  $f$  at points near  $\vec{x}_0$ . Similarly,  $f$  has a *local minimum* at  $\vec{x}_0$  if there is  $\delta > 0$ , such that

$$\|\vec{x} - \vec{x}_0\| < \delta \implies f(\vec{x}) \geq f(\vec{x}_0).$$

### Linear Approximation Condition

If  $\vec{x}_0$  is a local extreme, then it is a local extreme in all directions. So a necessary condition is that, if the directional derivative  $D_{\vec{v}}f(\vec{x}_0)$  exists, then  $D_{\vec{v}}f(\vec{x}_0) = 0$ . By taking  $\vec{v}$  to be the coordinate directions, we get the following.

**Proposition 8.6.1.** Suppose  $f(\vec{x})$  is defined near  $\vec{x}_0$  and has a local extreme at  $\vec{x}_0$ . If a partial derivative  $\frac{\partial f}{\partial x_i}(\vec{x}_0)$  exists, then the partial derivative must be zero.

For differentiable  $f$ , the condition can be expressed as  $\nabla f(=\vec{0})$ , or  $df = 0$ .

**Example 8.6.1.** The function  $f(x, y) = |x| + y$  satisfies  $f_y = 1 \neq 0$ . Therefore there is no local extreme, despite the fact that  $f_x$  may not exist.

**Example 8.6.2.** The function  $f(x, y) = \sqrt{|xy|}$  has the partial derivatives

$$f_x = \begin{cases} \frac{1}{2}\sqrt{\left|\frac{y}{x}\right|}, & \text{if } x > 0, \\ -\frac{1}{2}\sqrt{\left|\frac{y}{x}\right|}, & \text{if } x < 0, \\ 0, & \text{if } y = 0, \\ \text{does not exist,} & \text{if } x = 0, y \neq 0, \end{cases} \quad f_y = \begin{cases} \frac{1}{2}\sqrt{\left|\frac{x}{y}\right|}, & \text{if } y > 0, \\ -\frac{1}{2}\sqrt{\left|\frac{x}{y}\right|}, & \text{if } y < 0, \\ 0, & \text{if } x = 0, \\ \text{does not exist,} & \text{if } x \neq 0, y = 0. \end{cases}$$

Therefore the possible local extrema are  $(0, y_0)$  for any  $y_0$  and  $(x_0, 0)$  for any  $x_0$ . Since

$$f(x, y) = \sqrt{|xy|} \geq 0 = f(0, y_0) = f(x_0, 0),$$

the points on the two axes are indeed local minima.

**Example 8.6.3.** Consider the differentiable function  $z = z(x, y)$  given implicitly by  $x^2 - 2xy + 4yz + z^3 + 2y - z = 1$ . To find the possible local extrema of  $z(x, y)$ , we take the differential and get

$$(2x - 2y)dx + (-2x + 4z + 2)dy + (4y + 3z^2 - 1)dz = 0.$$

The possible local extrema are obtained by the condition  $dz = z_x dx + z_y dy = 0$  and the implicit equation. By solving

$$2x - 2y = 0, \quad -2x + 4z + 2 = 0, \quad x^2 - 2xy + 4yz + z^3 + 2y - z = 1,$$

we get the three possible local extrema  $(1, 1, 0)$ ,  $(-1, -1, -1)$ ,  $(-5, -5, -3)$  for  $z(x, y)$ .

**Example 8.6.4.** The continuous function  $f = x^2 - 2xy$  reaches its maximum and minimum on the compact subset  $|x| + |y| \leq 1$ . From  $f_x = 2x - 2y = 0$  and  $f_y = -2x = 0$ , we find the point  $(0, 0)$  to be the only possible local extreme in the interior  $|x| + |y| < 1$ . Then we look for the local extrema of  $f$  restricted to the boundary  $|x| + |y| = 1$ , which may be divided into four (open) line segments and four points.

On the segment  $x + y = 1$ ,  $0 < x < 1$ , we have  $f = x^2 - 2x(1 - x) = 3x^2 - 2x$ . From  $f_x = 6x - 2 = 0$  and  $x + y = 1$ , we find the possible local extreme  $\left(\frac{1}{3}, \frac{2}{3}\right)$  for  $f$  on the segment. Similarly, we find a possible local extreme  $\left(-\frac{1}{3}, -\frac{2}{3}\right)$  on the segment  $x + y = -1$ ,  $-1 < x < 0$ , and no possible local extrema on the segments  $x - y = 1$ ,  $0 < x < 1$  and  $-x + y = 1$ ,  $-1 < x < 0$ .

We also need to consider four points at the ends of the four segments, which are  $(1, 0)$ ,  $(-1, 0)$ ,  $(0, 1)$ ,  $(0, -1)$ . Comparing the values of  $f$  at all the possible local extrema

$$\begin{aligned} f(0, 0) &= 0, & f\left(\frac{1}{3}, \frac{2}{3}\right) &= f\left(-\frac{1}{3}, -\frac{1}{3}\right) = -\frac{1}{3}, \\ f(1, 0) &= f(-1, 0) = 1, & f(0, 1) &= f(0, -1) = 0, \end{aligned}$$

we find  $\pm\left(\frac{1}{3}, \frac{2}{3}\right)$  are the absolute minima and  $(\pm 1, 0)$  are the absolute maxima.

**Exercise 8.94.** Find possible local extrema.

1.  $x^2 y^3 (a - x - y)$ .
2.  $x + y + 4 \sin x \sin y$ .
3.  $xyz \log(x^2 + y^2 + z^2)$ .
4.  $x_1 x_2 \cdots x_n + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \cdots + \frac{a_n}{x_n}$ .
5.  $\frac{x_1 x_2 \cdots x_n}{(a + x_1)(x_1 + x_2) \cdots (x_n + b)}$ .

**Exercise 8.95.** Find possible local extrema of implicitly defined function  $z$ .

1.  $x^2 + y^2 + z^2 - 2x + 6z = 6$ .
2.  $(x^2 + y^2 + z^2)^2 = a^2(x^2 + y^2 - z^2)$ .
3.  $x^3 + y^3 + z^3 - xyz = 1$ .

**Exercise 8.96.** Find the absolute extrema for the functions on the given domain.

1.  $x + y + z$  on  $\{(x, y, z): x^2 + y^2 \leq z \leq 1\}$ .
2.  $\sin x + \sin y - \sin(x + y)$  on  $\{(x, y): x \geq 0, y \geq 0, x + y \leq 2\pi\}$ .

**Exercise 8.97.** What is the shortest distance between two straight lines?

**Exercise 8.98.** Suppose  $f$  is continuous near  $\vec{0}$  and is differentiable near (but not at)  $\vec{0}$ . Prove that if  $\nabla f(\vec{v}) \cdot \vec{v} \geq 0$  for any  $\vec{v}$  near  $\vec{0}$ , then  $\vec{0}$  is a local minimum.

## Quadratic Approximation Condition

The linear approximation may be used to find the potential local extrema. The quadratic approximation may be further used to determine whether the candidates are indeed local extrema.

If a function is second order differentiable, then the second order derivative is the *Hessian* of the function. Under the assumption of Theorem 8.5.1, the Hessian



may be computed by the second order partial derivatives

$$\begin{aligned} h_f(\vec{v}) &= \sum_{1 \leq i \leq n} \frac{\partial^2 f}{\partial x_i^2} v_i^2 + 2 \sum_{1 \leq i < j \leq n} \frac{\partial^2 f}{\partial x_i \partial x_j} v_i v_j \\ &= \frac{\partial^2 f}{\partial x_1^2} v_1^2 + \frac{\partial^2 f}{\partial x_2^2} v_2^2 + \cdots + \frac{\partial^2 f}{\partial x_n^2} v_n^2 \\ &\quad + 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} v_1 v_2 + 2 \frac{\partial^2 f}{\partial x_1 \partial x_3} v_1 v_3 + \cdots + 2 \frac{\partial^2 f}{\partial x_{n-1} \partial x_n} v_{n-1} v_n. \end{aligned}$$

**Proposition 8.6.2.** *Suppose  $f(\vec{x})$  is second order differentiable at  $\vec{x}_0$ , and the first order derivative vanishes at  $\vec{x}_0$ .*

1. *If the Hessian is positive definite, then  $\vec{x}_0$  is a local minimum.*
2. *If the Hessian is negative definite, then  $\vec{x}_0$  is a local maximum.*
3. *If the Hessian is indefinite, then  $\vec{x}_0$  is not a local extreme.*

*Proof.* The assumption says that  $f$  is approximated by  $f(\vec{x}_0) + \frac{1}{2} h_f(\Delta \vec{x})$  near  $\vec{x}_0$ . This means that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|\Delta \vec{x}\| < \delta \implies \left| f(\vec{x}) - f(\vec{x}_0) - \frac{1}{2} h_f(\Delta \vec{x}) \right| \leq \epsilon \|\Delta \vec{x}\|^2.$$

The Hessian  $h_f(\vec{v})$  is a continuous function and reaches its maximum and minimum on the unit sphere  $\{\vec{v}: \|\vec{v}\| = 1\}$ , which is a compact subset. Suppose the Hessian is positive definite. Then the minimum on the subset is reached at some point  $\vec{v}_0$ , with value  $m = h_f(\vec{v}_0) > 0$ . This means that  $h_f(\vec{v}) \geq m > 0$  for any  $\vec{v}$  of unit length. By  $h_f(c\vec{v}) = c^2 h_f(\vec{v})$ , we conclude that  $h_f(\vec{v}) \geq m \|\vec{v}\|^2$  for any  $\vec{v}$ . See Exercise 6.97.

Fix  $\epsilon > 0$  satisfying  $\epsilon < \frac{m}{2}$ . Then there is  $\delta > 0$ , such that

$$\begin{aligned} \|\Delta \vec{x}\| < \delta &\implies \left| f(\vec{x}) - f(\vec{x}_0) - \frac{1}{2} h_f(\Delta \vec{x}) \right| \leq \epsilon \|\Delta \vec{x}\|^2 \\ &\implies f(\vec{x}) - f(\vec{x}_0) \geq \frac{1}{2} h_f(\Delta \vec{x}) - \epsilon \|\Delta \vec{x}\|^2 \geq \left( \frac{m}{2} - \epsilon \right) \|\Delta \vec{x}\|^2 \\ &\implies f(\vec{x}) - f(\vec{x}_0) > 0 \text{ for } \Delta \vec{x} \neq \vec{0}. \end{aligned}$$

This proves that  $\vec{x}_0$  is a *strict* local minimum.

By similar reason, if the Hessian is negative definite, then  $\vec{x}_0$  is a *strict* local maximum.

Finally, suppose the Hessian is indefinite. Then we have  $h_f(\vec{v}) > 0$  and  $h_f(\vec{w}) < 0$  for some vectors  $\vec{v}$  and  $\vec{w}$ . The restriction of  $f$  on the straight line  $\vec{x} = \vec{x}_0 + t\vec{v}$  has Hessian  $h_f(t\vec{v}) = t^2 h_f(\vec{v})$ , which is positive definite. Therefore  $\vec{x}_0$  is a strict minimum of the restriction. Similarly, by  $h_f(\vec{w}) < 0$ , the restriction of  $f$  on the straight line  $\vec{x} = \vec{x}_0 + t\vec{w}$  has  $\vec{x}_0$  as a strict local maximum. Combining the two facts,  $\vec{x}_0$  is not a local extreme of  $f$ .  $\square$

**Example 8.6.5.** We try to find the local extrema of  $f = x^3 + y^2 + z^2 + 12xy + 2z$ . By solving

$$f_x = 3x^2 + 12y = 0, \quad f_y = 2y + 12x = 0, \quad f_z = 2z + 2 = 0,$$

we find two possible local extrema  $\vec{a} = (0, 0, -1)$  and  $\vec{b} = (24, -144, -1)$ . The Hessian of  $f$  at the two points are

$$h_{f,\vec{a}}(u, v, w) = 2v^2 + 2w^2 + 24uv, \quad h_{f,\vec{b}}(u, v, w) = 144u^2 + 2v^2 + 2w^2 + 24uv.$$

Since  $h_{f,\vec{a}}(1, 1, 0) = 26 > 0$ ,  $h_{f,\vec{a}}(-1, 1, 0) = -22 < 0$ ,  $\vec{a}$  is not a local extreme. Since  $h_{f,\vec{b}} = (12u + v)^2 + v^2 + 2w^2 > 0$  for  $(u, v, w) \neq \vec{0}$ ,  $\vec{b}$  is a local minimum.

**Example 8.6.6.** We study whether the three possibilities in Example 8.6.3 are indeed local extrema. By  $z_x = \frac{-2x + 2y}{4y + 3z^2 - 1}$ ,  $z_y = \frac{2x - 4z - 2}{4y + 3z^2 - 1}$  and the fact that  $z_x = z_y = 0$  at the three points, we have

$$z_{xx} = \frac{-2}{4y + 3z^2 - 1}, \quad z_{xy} = \frac{2}{4y + 3z^2 - 1}, \quad z_{yy} = \frac{-8(x - 2z - 1)}{(4y + 3z^2 - 1)^2},$$

at the three points. Then we get the Hessians at the three points

$$\begin{aligned} h_{z,(1,1,0)}(u, v) &= -\frac{2}{3}u^2 + \frac{4}{3}uv, \\ h_{z,(-1,-1,-1)}(u, v) &= u^2 - 2uv, \\ h_{z,(-5,-5,-3)}(u, v) &= -\frac{1}{3}u^2 + \frac{2}{3}uv. \end{aligned}$$

By taking  $(u, v) = (1, 1)$  and  $(1, -1)$ , we see the Hessians are indefinite. Thus none of the three points are local extrema.

**Example 8.6.7.** The only possible local extreme for the function  $f(x, y) = x^3 + y^2$  is  $(0, 0)$ , where the Hessian  $h_f(u, v) = 2v^2$ . Although the Hessian is non-negative, it is not positive definite. In fact, the Hessian is positive semi-definite, and  $(0, 0)$  is not a local extreme.

The problem is that the restriction to the  $x$ -axis is  $f(x, 0) = x^3$ , and the local extreme problem cannot be solved by the quadratic approximation. To study the local extreme of  $f(x, y) = x^3 + y^2$  at  $(0, 0)$ , we must consider the quadratic approximation in  $y$ -direction and the cubic approximation in  $x$ -direction.

**Exercise 8.99.** Try your best to determine whether the possible local extrema in Exercises 8.94 and 8.95 are indeed local maxima or local minima.

**Exercise 8.100.** Suppose a two variable function  $f(x, y)$  has continuous second order partial derivatives at  $(x_0, y_0)$ . Suppose  $f_x = f_y = 0$  at  $(x_0, y_0)$ .

1. Prove that if  $f_{xx} > 0$  and  $f_{xx}f_{yy} - f_{xy}^2 > 0$ , then  $(x_0, y_0)$  is a local minimum.
2. Prove that if  $f_{xx} < 0$  and  $f_{xx}f_{yy} - f_{xy}^2 > 0$ , then  $(x_0, y_0)$  is a local maximum.
3. Prove that if  $f_{xx} \neq 0$  and  $f_{xx}f_{yy} - f_{xy}^2 < 0$ , then  $(x_0, y_0)$  is not a local extreme.

**Exercise 8.101.** Show that the function

$$f(x, y) = \begin{cases} x^2 + y^2 + \frac{x^3 y^3}{(x^2 + y^2)^3}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0), \end{cases}$$

has first and second order partial derivatives and satisfy  $\nabla f(0, 0) = 0$  and  $h_f(u, v) > 0$  for  $(u, v) \neq (0, 0)$ . However, the function does not have a local extreme at  $(0, 0)$ .

The counterexample is not continuous at  $(0, 0)$ . Can you find a counterexample that is differentiable at  $(0, 0)$ ?

**Exercise 8.102.** Suppose a function  $f(\vec{x})$  has continuous third order partial derivatives at  $\vec{x}_0$ , such that all the first and second order partial derivatives vanish at  $\vec{x}_0$ . Prove that if some third order partial derivative is nonzero, then  $\vec{x}_0$  is not a local extreme.

**Exercise 8.103.** Let  $k$  be a natural number. Show that the equations

$$x^k - z + \sin(y^k + w) = 0, \quad e^{x^k - w} + y^k - z = 1$$

implicitly define  $z$  and  $w$  as differentiable functions of  $x$  and  $y$  near  $(x, y, z, w) = (0, 0, 0, 0)$ . Moreover, determine whether  $(0, 0)$  is a local extreme of  $z = z(x, y)$ .

Example 8.6.7 shows that, when the Hessian is in none of the three cases in Proposition 8.6.2, the quadratic approximation may not be enough for concluding the local extreme. In the single variable case, this problem also arises and can be solved by considering cubic or even higher order approximation. However, the multivariable case does not have similar solution due to two problems. The first is that we may need approximations of different orders in different directions, so that a neat criterion like Proposition 3.5.1 is impossible. The second is that even if we can use approximations of the same order in all directions, there is no general technique such as completing the squares, from which we can determine the positive or negative definiteness.

### Constrained Extreme: Linear Approximation Condition

Suppose  $G: \mathbb{R}^n \rightarrow \mathbb{R}^k$  is a map and  $G(\vec{x}_0) = \vec{c}$ . A function  $f(\vec{x})$  has a *local maximum at  $\vec{x}_0$  under the constraint  $G(\vec{x}) = \vec{c}$*  if there is  $\delta > 0$ , such that

$$G(\vec{x}) = \vec{c}, \|\vec{x} - \vec{x}_0\| < \delta \implies f(\vec{x}) \leq f(\vec{x}_0).$$

In other words, the value of  $f$  at  $\vec{x}_0$  is biggest among the values of  $f$  at points near  $\vec{x}_0$  and satisfying  $G(\vec{x}) = \vec{c}$ . Similarly,  $f$  has a *local minimum at  $\vec{x}_0$  under the constraint  $G(\vec{x}) = \vec{c}$*  if there is  $\delta > 0$ , such that

$$G(\vec{x}) = \vec{c}, \|\vec{x} - \vec{x}_0\| < \delta \implies f(\vec{x}) \geq f(\vec{x}_0).$$

For example, the hottest and the coldest places in the world are the extrema of the temperature function  $T(x, y, z)$  under the constraint  $g(x, y, z) = x^2 + y^2 + z^2 = R^2$ , where  $R$  is the radius of the earth.

Suppose  $G$  is continuously differentiable and  $\vec{c}$  is a regular value. Then by Proposition 8.4.3,  $G(\vec{x}) = \vec{c}$  defines an  $(n - k)$ -dimensional “constraint submanifold”  $M \subset \mathbb{R}^n$ . A local extreme of  $f$  under constraint  $G = \vec{c}$  simply means a local extreme of the restriction  $f|_M$  of  $f$  to  $M$ . If  $\vec{x}_0 \in M$  is such a local extreme, then it is a local extreme for the restriction of  $f$  on any curve in  $M$  passing through  $\vec{x}_0$ . Let

$\phi(t)$  be such a curve, with  $\phi(0) = \vec{x}_0$ . Then 0 is a local extreme for  $f(\phi(t))$ . For differentiable  $f$  and  $\phi$ , by Proposition 3.3.1, we have

$$0 = \left. \frac{d}{dt} \right|_{t=0} f(\phi(t)) = f'(\vec{x}_0)(\phi'(0)).$$

By (8.4.2), all the vectors  $\phi'(0)$  for all possible  $\phi$  form the tangent space  $T_{\vec{x}_0}M = \{\vec{v}: G'(\vec{x}_0)(\vec{v}) = \vec{0}\}$ , which is the kernel of the linear transform  $G'(\vec{x}_0)$ . Therefore we get the necessary condition

$$G'(\vec{x}_0)(\vec{v}) = \vec{0} \implies f'(\vec{x}_0)(\vec{v}) = 0,$$

for the local extreme under constraint.

To calculate the necessary condition, suppose  $G = (g_1, \dots, g_k)$ . Then

$$G'(\vec{x}_0)(\vec{v}) = (\nabla g_1(\vec{x}_0) \cdot \vec{v}, \dots, \nabla g_k(\vec{x}_0) \cdot \vec{v}), \quad f'(\vec{x}_0)(\vec{v}) = \nabla f(\vec{x}_0) \cdot \vec{v}.$$

Therefore the necessary condition becomes

$$\nabla g_1(\vec{x}_0) \cdot \vec{v} = \dots = \nabla g_k(\vec{x}_0) \cdot \vec{v} = 0 \implies \nabla f(\vec{x}_0) \cdot \vec{v} = 0.$$

Linear algebra tells us that this is equivalent to  $\nabla f(\vec{x}_0)$  being a linear combination of  $\nabla g_1(\vec{x}_0), \dots, \nabla g_k(\vec{x}_0)$ .

**Proposition 8.6.3.** *Suppose  $G = (g_1, \dots, g_k): \mathbb{R}^n \rightarrow \mathbb{R}^k$  is a continuously differentiable map near  $\vec{x}_0$ , and  $G(\vec{x}_0) = \vec{c}$  is a regular value of  $G$ . Suppose a function  $f$  defined near  $\vec{x}_0$  is differentiable at  $\vec{x}_0$ . If  $f$  has a local extreme at  $\vec{x}_0$  under the constraint  $G(\vec{x}) = \vec{c}$ , then  $\nabla f(\vec{x}_0)$  is a linear combination of  $\nabla g_1(\vec{x}_0), \dots, \nabla g_k(\vec{x}_0)$ :*

$$\nabla f(\vec{x}_0) = \lambda_1 \nabla g_1(\vec{x}_0) + \dots + \lambda_k \nabla g_k(\vec{x}_0).$$

The coefficients  $\lambda_1, \dots, \lambda_k$  are called the *Lagrange multipliers*. The condition can also be written as an equality of linear functionals

$$f'(\vec{x}_0) = \vec{\lambda} \cdot G'(\vec{x}_0).$$

**Example 8.6.8.** The local extrema problem in Example 8.6.3 can also be considered as the local extrema of the function  $f(x, y, z) = z$  under the constraint  $g(x, y, z) = x^2 - 2xy + 4yz + z^3 + 2y - z = 1$ . At local extrema, we have

$$\nabla f = (0, 0, 1) = \lambda \nabla g = \lambda(2x - 2y, -2x + 4z + 2, 4y + 2z^2 - 1).$$

This is the same as

$$0 = \lambda(2x - 2y), \quad 0 = \lambda(-2x + 4z + 2), \quad 1 = \lambda(4y + 2z^2 - 1).$$

The third equality tells us  $\lambda \neq 0$ . Thus the first two equalities become  $2x - 2y = 0$  and  $-2x + 4z + 2 = 0$ . These are the conditions we found in Example 8.6.3.

**Example 8.6.9.** We try to find the possible local extrema of  $f = xy + yz + zx$  on the sphere  $x^2 + y^2 + z^2 = 1$ . By  $\nabla f = (y + z, z + x, x + y)$  and  $\nabla(x^2 + y^2 + z^2) = (2x, 2y, 2z)$ , we have

$$y + z = 2\lambda x, \quad z + x = 2\lambda y, \quad x + y = 2\lambda z, \quad x^2 + y^2 + z^2 = 1$$

at local extrema. Adding the first three equalities together, we get  $(\lambda - 1)(x + y + z) = 0$ .

If  $\lambda = 1$ , then  $x = y = z$  from the first three equations. Substituting into the fourth equation, we get  $3x^2 = 1$  and two possible extrema  $\pm \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$ .

If  $\lambda \neq 1$ , then  $x + y + z = 0$  and the first three equations become  $(2\lambda + 1)x = (2\lambda + 1)y = (2\lambda + 1)z = 0$ . Since  $(0, 0, 0)$  does not satisfy  $x^2 + y^2 + z^2 = 1$ , we must have  $2\lambda + 1 = 0$ , and the four equations is equivalent to  $x + y + z = 0$  and  $x^2 + y^2 + z^2 = 1$ , which is a circle in  $\mathbb{R}^3$ .

Thus the possible local extrema are  $\pm \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$  and the points on the circle  $x + y + z = 0$ ,  $x^2 + y^2 + z^2 = 1$ .

Note that the sphere is compact and the continuous function  $f$  reaches its maximum and minimum on the sphere. Since  $f = 1$  at  $\pm \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$  and  $f = \frac{1}{2}[(x + y + z)^2 - (x^2 + y^2 + z^2)] = -\frac{1}{2}$  along the circle  $x + y + z = 0$ ,  $x^2 + y^2 + z^2 = 1$ , we find the maximum is 1 and the minimum is  $-\frac{1}{2}$ .

We can also use  $f_x = y + z = 0$ ,  $f_y = z + x = 0$ ,  $f_z = x + y = 0$  to find the possible local extreme  $(0, 0, 0)$  of  $f$  in the interior  $x^2 + y^2 + z^2 < 1$  of the ball  $x^2 + y^2 + z^2 \leq 1$ . By comparing  $f(0, 0, 0) = 0$  with the maximum and minimum of  $f$  on the sphere, we find  $f$  reaches its extrema on the ball at boundary points.

**Example 8.6.10.** We try to find the possible local extrema of  $f = xyz$  on the circle given by  $g_1 = x + y + z = 0$  and  $g_2 = x^2 + y^2 + z^2 = 6$ . The necessary condition is

$$\nabla f = (yz, zx, xy) = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 = (\lambda_1 + 2\lambda_2 x, \lambda_1 + 2\lambda_2 y, \lambda_1 + 2\lambda_2 z).$$

Canceling  $\lambda_1$  from the three equations, we get

$$(x - y)(z + 2\lambda_2) = 0, \quad (x - z)(y + 2\lambda_2) = 0.$$

If  $x \neq y$  and  $x \neq z$ , then  $y = z = -2\lambda_2$ . Therefore at least two of  $x, y, z$  are equal.

If  $x = y$ , then the two constraints  $g_1 = 0$  and  $g_2 = 6$  tell us  $2x + z = 0$  and  $2x^2 + z^2 = 6$ . Canceling  $z$ , we get  $6x^2 = 6$  and two possible local extrema  $(1, 1, -2)$  and  $(-1, -1, 2)$ . By assuming  $x = z$  or  $y = z$ , we get four more possible local extrema  $(1, -2, 1)$ ,  $(-1, 2, -1)$ ,  $(-2, 1, 1)$ ,  $(2, -1, -1)$ .

By evaluating the function at the six possible local extrema, we find the absolute maximum 2 is reached at three points and the absolute minimum  $-2$  is reached at the other three points.

**Exercise 8.104.** Find possible local extrema under the constraint.

1.  $x^p y^q z^r$  under the constraint  $x + y + z = a$ ,  $x, y, z > 0$ , where  $p, q, r > 0$ .
2.  $\sin x \sin y \sin z$  under the constraint  $x + y + z = \frac{\pi}{2}$ .
3.  $x_1 + x_2 + \cdots + x_n$  under the constraint  $x_1 x_2 \cdots x_n = a$ .
4.  $x_1 x_2 \cdots x_n$  under the constraint  $x_1 + x_2 + \cdots + x_n = a$ .

5.  $x_1 x_2 \cdots x_n$  under the constraint  $x_1 + x_2 + \cdots + x_n = a$ ,  $x_1^2 + x_2^2 + \cdots + x_n^2 = b$ .
6.  $x_1^p + x_2^p + \cdots + x_n^p$  under the constraint  $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$ , where  $p > 0$ .

Exercise 8.105. Prove inequalities.

1.  $\sqrt[n]{x_1 x_2 \cdots x_n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n}$  for  $x_i \geq 0$ .
2.  $\frac{x_1^p + x_2^p + \cdots + x_n^p}{n} \geq \left( \frac{x_1 + x_2 + \cdots + x_n}{n} \right)^p$  for  $p \geq 1$ ,  $x_i \geq 0$ . What if  $0 < p < 1$ ?

Exercise 8.106. Derive Hölder inequality in Exercise 3.75 by considering the function  $\sum b_i^q$  of  $(b_1, b_2, \dots, b_n)$  under the constraint  $\sum a_i b_i = 1$ .

Exercise 8.107. Suppose  $A$  is a symmetric matrix. Prove that if the quadratic form  $q(\vec{x}) = A\vec{x} \cdot \vec{x}$  reaches maximum or minimum at  $\vec{v}$  on the unit sphere  $\{\vec{x}: \|\vec{x}\|_2 = 1\}$ , then  $\vec{v}$  is an eigenvector of  $A$ .

Exercise 8.108. Fix the base triangle and the height of a pyramid. When is the total area of the side faces smallest?

Exercise 8.109. The intersection of the plane  $x + y - z = 0$  and the ellipsoid  $x^2 + y^2 + z^2 - xy - yz - zx = 1$  is an ellipse centered at the origin. Find the lengths of the two axes of the ellipse.

## Constrained Extreme: Quadratic Approximation Condition

After finding the possible local extrema by using the linear approximation, we may further use the quadratic approximation to determine whether the candidates are indeed local extrema.

Assume we are in the situation described in Proposition 8.6.3. Further assume that  $f$  and  $G$  have continuous second order partial derivatives. Then  $f$  and  $g_i$  have quadratic approximations near the possible local extreme  $\vec{x}_0$

$$p_f(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot \Delta\vec{x} + \frac{1}{2}h_f(\Delta\vec{x}),$$

$$p_{g_i}(\vec{x}) = g_i(\vec{x}_0) + \nabla g_i(\vec{x}_0) \cdot \Delta\vec{x} + \frac{1}{2}h_{g_i}(\Delta\vec{x}).$$

The original constrained local maximum problem

$$f(\vec{x}) < f(\vec{x}_0) \text{ for } \vec{x} \text{ near } \vec{x}_0, \vec{x} \neq \vec{x}_0, \text{ and } g_i(\vec{x}) = c_i, 1 \leq i \leq k$$

is approximated by the similar constrained problem

$$p_f(\vec{x}) < p_f(\vec{x}_0) = f(\vec{x}_0) \text{ for } \vec{x} \text{ near } \vec{x}_0, \vec{x} \neq \vec{x}_0, \text{ and } p_{g_i}(\vec{x}) = c_i, 1 \leq i \leq k.$$

Here the constraint is also quadratically approximated.

The approximate constrained problem is the same as the following ( $\vec{v} = \Delta \vec{x}$ )

$$\begin{aligned} \nabla f(\vec{x}_0) \cdot \vec{v} + \frac{1}{2} h_f(\vec{v}) &< 0 \\ \text{for small } \vec{v} \neq \vec{0} \text{ satisfying } \nabla g_i(\vec{x}_0) \cdot \vec{v} + \frac{1}{2} h_{g_i}(\vec{v}) &= 0, \quad 1 \leq i \leq k. \end{aligned}$$

The problem is not easy to solve because both approximate function and approximate constraints are mixes of linear and quadratic terms. To remove the mixture, we use the Lagrange multipliers  $\lambda_i$  appearing in Proposition 8.6.3 to introduce

$$\tilde{f}(\vec{x}) = f(\vec{x}) - \vec{\lambda} \cdot G(\vec{x}) = f(\vec{x}) - \lambda_1 g_1(\vec{x}) - \cdots - \lambda_k g_k(\vec{x}).$$

On the constraint submanifold  $M$ , we have  $\tilde{f}|_M = f|_M - \vec{\lambda} \cdot \vec{c}$ . Therefore the two restricted functions  $\tilde{f}|_M$  and  $f|_M$  differ by a constant, and they have the same constrained extreme. This means that the problem can be reduced to

$$\begin{aligned} \nabla \tilde{f}(\vec{x}_0) \cdot \vec{v} + \frac{1}{2} h_{\tilde{f}}(\vec{v}) &< 0 \\ \text{for small } \vec{v} \neq \vec{0} \text{ satisfying } \nabla g_i(\vec{x}_0) \cdot \vec{v} + \frac{1}{2} h_{g_i}(\vec{v}) &= 0, \quad 1 \leq i \leq k. \end{aligned}$$

Since

$$\begin{aligned} \nabla \tilde{f}(\vec{x}_0) &= \nabla f(\vec{x}_0) - \lambda_1 \nabla g_1(\vec{x}_0) - \cdots - \lambda_k \nabla g_k(\vec{x}_0) = 0, \\ h_{\tilde{f}} &= h_f - \lambda_1 h_{g_1} - \cdots - \lambda_k h_{g_k}, \end{aligned}$$

The modified problem is actually

$$h_{\tilde{f}}(\vec{v}) < 0 \text{ for small } \vec{v} \neq \vec{0} \text{ satisfying } \nabla g_i(\vec{x}_0) \cdot \vec{v} + \frac{1}{2} h_{g_i}(\vec{v}) = 0, \quad 1 \leq i \leq k.$$

The constraint is linearly approximated by the solution of  $\nabla g_i(\vec{x}_0) \cdot \vec{v} = 0$ . By Exercise 8.33, the difference between the linear approximation and the actual constraint changes  $h_{\tilde{f}}(\vec{v})$  by an amount of  $o(\|\vec{v}\|^2)$ . In particular, this difference does not affect the negativity of  $h_{\tilde{f}}(\vec{v})$  for small  $\vec{v}$ , and the modified problem is therefore the same as

$$h_{\tilde{f}}(\vec{v}) < 0 \text{ for small } \vec{v} \neq \vec{0} \text{ satisfying } \nabla g_i(\vec{x}_0) \cdot \vec{v} = 0, \quad 1 \leq i \leq k.$$

**Proposition 8.6.4.** *Suppose  $G = (g_1, \dots, g_k): \mathbb{R}^n \rightarrow \mathbb{R}^k$  has continuous first order partial derivatives near  $\vec{x}_0$ , and  $G(\vec{x}_0) = \vec{c}$  is a regular value of  $G$ . Suppose a function  $f$  is second order differentiable at  $\vec{x}_0$ , and  $f'(\vec{x}_0) = \vec{\lambda} \cdot G'(\vec{x}_0)$  for a vector  $\vec{\lambda} \in \mathbb{R}^k$ . Denote the quadratic form*

$$q = h_f - \vec{\lambda} \cdot h_G = h_f - \lambda_1 h_{g_1} - \cdots - \lambda_k h_{g_k}.$$

1. *If  $q(\vec{v})$  is positive definite for vectors  $\vec{v}$  satisfying  $G'(\vec{x}_0)(\vec{v}) = \vec{0}$ , then  $\vec{x}_0$  is a local minimum of  $f$  under the constraint  $G(\vec{x}) = \vec{c}$ .*

2. If  $q(\vec{v})$  is negative definite for vectors  $\vec{v}$  satisfying  $G'(\vec{x}_0)(\vec{v}) = \vec{0}$ , then  $\vec{x}_0$  is a local maximum of  $f$  under the constraint  $G(\vec{x}) = \vec{c}$ .
3. If  $q(\vec{v})$  is indefinite for vectors  $\vec{v}$  satisfying  $G'(\vec{x}_0)(\vec{v}) = \vec{0}$ , then  $\vec{x}_0$  is not a local extreme of  $f$  under the constraint  $G(\vec{x}) = \vec{c}$ .

*Proof.* Since  $G(\vec{x}) = \vec{c}$  implies  $\tilde{f}(\vec{x}) = f(\vec{x}) - \vec{\lambda} \cdot G(\vec{x}) = f(\vec{x}) - \vec{\lambda} \cdot \vec{c}$ , the local extreme problem for  $f(\vec{x})$  under the constraint  $G(\vec{x}) = \vec{c}$  is the same as the local extreme problem for  $\tilde{f}(\vec{x})$  under the constraint  $G(\vec{x}) = \vec{c}$ .

Since  $\vec{c}$  is a regular value, by Proposition 8.4.3, the constraint  $G(\vec{x}) = \vec{c}$  means that, near  $\vec{x}_0$ , some choice of  $k$  coordinates of  $\vec{x}$  can be written as a continuously differentiable map of the other  $n - k$  coordinates  $\vec{y}$ . Thus the solution of the constraint is a map  $\vec{x} = H(\vec{y})$  of some coordinates  $\vec{y}$  of  $\vec{x}$ . In fact,  $\vec{x} = H(\vec{y})$  is a regular parameterization of the constraint submanifold. See Proposition 8.4.2. Then the problem is to determine whether  $\vec{y}_0$  is an unconstrained local extreme of  $\tilde{f}(H(\vec{y}))$ . The problem can be solved by applying Proposition 8.6.2 to  $\tilde{f}(H(\vec{y}))$ .

Since  $\vec{x} = H(\vec{y})$  is the solution of the constraint  $G(\vec{x}) = \vec{c}$ , the linear approximation  $\vec{x} = \vec{x}_0 + H'(\vec{y}_0)(\Delta\vec{y})$  is the solution to the linear approximation of the constraint  $G'(\vec{x}_0)(\Delta\vec{x}) = \vec{0}$ . In other words, vectors of the form  $\vec{v} = H'(\vec{y}_0)(\vec{u})$  (i.e., the image of the linear transform  $H'(\vec{y}_0)$ ) are exactly the vectors satisfying  $G'(\vec{x}_0)(\vec{v}) = \vec{0}$ . See the discussion before the Implicit Function Theorem (Theorem 8.3.2) for a more explicit explanation of the fact.

By the computation before the proposition, the quadratic approximation of  $\tilde{f}(\vec{x})$  near  $\vec{x}_0$  is

$$\tilde{f}(\vec{x}) = \tilde{f}(\vec{x}_0) + \frac{1}{2}q(\Delta\vec{x}) + R_2(\Delta\vec{x}), \quad R_2(\Delta\vec{x}) = o(\|\Delta\vec{x}\|^2).$$

The linear approximation of  $\vec{x} = H(\vec{y})$  near  $\vec{y}_0$  is

$$\vec{x} = H(\vec{y}) = \vec{x}_0 + H'(\vec{y}_0)(\Delta\vec{y}) + R_1(\Delta\vec{y}), \quad R_1(\Delta\vec{y}) = o(\|\Delta\vec{y}\|).$$

By the second part of Exercise 8.33, substituting this into the quadratic approximation of  $\tilde{f}$  gives

$$\tilde{f}(H(\vec{y})) = \tilde{f}(\vec{x}_0) + \frac{1}{2}q(H'(\vec{y}_0)(\Delta\vec{y})) + R(\Delta\vec{y}), \quad R(\Delta\vec{y}) = o(\|\Delta\vec{y}\|^2).$$

By Proposition 8.6.2, the nature of the quadratic form  $q(H'(\vec{y}_0)(\vec{u}))$  determines the nature of local extreme at  $\vec{y}_0$ . By the earlier discussion, the quadratic form  $q(H'(\vec{y}_0)(\vec{u}))$  is exactly the restriction of the quadratic form  $q(\vec{v})$  to vectors  $\vec{v} = H'(\vec{y}_0)(\vec{u})$  satisfying  $G'(\vec{x}_0)(\vec{v}) = \vec{0}$ .  $\square$

We note that the proposition does not require  $G$  to be second order differentiable. Moreover,  $f$  is only required to be second order differentiable, and is not required to have partial derivatives.

We also note that the proof only makes use of some regular parameterization  $H(\vec{y})$  of the constraint submanifold, and does not make use of the specific knowledge that  $\vec{y}$  is some choice of coordinates.



**Example 8.6.11.** We try to find the possible local extrema of  $f = xy^2$  on the circle  $g = x^2 + y^2 = 3$ . The linear condition  $\nabla f = (y^2, 2xy) = \lambda \nabla g = \lambda(2x, 2y)$  gives

$$y^2 = 2\lambda x, \quad 2xy = 2\lambda y, \quad x^2 + y^2 = 3.$$

The solutions are  $\lambda = 0$  and  $(x, y) = (\pm\sqrt{3}, 0)$ , or  $\lambda \neq 0$  and  $(x, y) = (\pm 1, \pm\sqrt{2})$  (the two  $\pm$  are independent, so that there are four choices).

To determine whether they are indeed local extrema, we compute the Hessians  $h_f(u, v) = 4yuv + 2xv^2$  and  $h_g(u, v) = 2u^2 + 2v^2$ . Together with  $\nabla g = (2x, 2y)$ , we have the following table showing what happens at the six points.

$(x_0, y_0)$	$\lambda$	$h_f - \lambda h_g$	$\nabla g \cdot (u, v) = 0$	restricted $h_f - \lambda h_g$
$(\sqrt{3}, 0)$	0	$2\sqrt{3}v^2$	$2\sqrt{3}u = 0$	$2\sqrt{3}v^2$
$(-\sqrt{3}, 0)$	0	$-2\sqrt{3}v^2$	$-2\sqrt{3}u = 0$	$-2\sqrt{3}v^2$
$(1, \sqrt{2})$	1	$-2u^2 + 4\sqrt{2}uv$	$2u + 2\sqrt{2}v = 0$	$-6u^2$
$(1, -\sqrt{2})$	1	$-2u^2 - 4\sqrt{2}uv$	$2u - 2\sqrt{2}v = 0$	$-6u^2$
$(-1, \sqrt{2})$	-1	$2u^2 + 4\sqrt{2}uv$	$-2u + 2\sqrt{2}v = 0$	$6u^2$
$(-1, -\sqrt{2})$	-1	$2u^2 - 4\sqrt{2}uv$	$-2u - 2\sqrt{2}v = 0$	$6u^2$

We note that the unrestricted  $h_f - \lambda h_g$  is a two variable quadratic form, and the restricted  $h_f - \lambda h_g$  is a single variable quadratic form. So the form  $2\sqrt{3}v^2$  in the third column is, as a two variable form, not positive definite (only positive semi-definite). On the other hand, the same form  $2\sqrt{3}v^2$  in the fifth column is, as a single variable form, positive definite.

We conclude that  $(1, \pm\sqrt{2})$  and  $(-\sqrt{3}, 0)$  are local maxima, and  $(-1, \pm\sqrt{2})$  and  $(\sqrt{3}, 0)$  are local minima.

**Example 8.6.12.** In Example 8.6.9, we found that  $\pm \left( \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)$  are the possible extrema of  $f = xy + yz + zx$  on the sphere  $x^2 + y^2 + z^2 = 1$ . To determine whether they are indeed local extrema, we compute the Hessians  $h_f(u, v, w) = 2uv + 2vw + 2wu$  and  $h_{x^2+y^2+z^2}(u, v, w) = 2u^2 + 2v^2 + 2w^2$ . At the two points, we have  $\lambda = 1$  and

$$h_f - \lambda h_{x^2+y^2+z^2} = -2u^2 - 2v^2 - 2w^2 + 2uv + 2vw + 2wu.$$

By  $\nabla(x^2 + y^2 + z^2) = (2x, 2y, 2z)$ , we need to consider the sign of  $h_f - \lambda h_{x^2+y^2+z^2}$  for those  $(u, v, w)$  satisfying

$$\pm \left( \frac{1}{\sqrt{3}}u + \frac{1}{\sqrt{3}}v + \frac{1}{\sqrt{3}}w \right) = 0.$$

Substituting the solution  $w = -u - v$ , we get

$$h_f - \lambda h_{x^2+y^2+z^2} = -6u^2 - 6v^2 - 6uv = -6 \left( u + \frac{1}{2}v \right)^2 - \frac{9}{2}v^2,$$

which is negative definite. Thus  $\pm \left( \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)$  are local maxima.

**Exercise 8.110.** Determine whether the possible local extrema in Exercise 8.104 are indeed local maxima or local minima.

## 8.7 Additional Exercise

### Orthogonal Change of Variable

A change of variable  $\vec{x} = F(\vec{y}): \mathbb{R}^n \rightarrow \mathbb{R}^n$  is *orthogonal* if the vectors

$$\vec{x}_{y_1} = \frac{\partial \vec{x}}{\partial y_1}, \quad \vec{x}_{y_2} = \frac{\partial \vec{x}}{\partial y_2}, \quad \dots, \quad \vec{x}_{y_n} = \frac{\partial \vec{x}}{\partial y_n}$$

are orthogonal.

Exercise 8.111. Prove that an orthogonal change of variable satisfies  $\frac{\partial y_i}{\partial x_j} = \frac{1}{\|\vec{x}_{y_i}\|_2^2} \frac{\partial x_j}{\partial y_i}$ .

Exercise 8.112. Is the inverse  $\vec{y} = F^{-1}(\vec{x})$  of an orthogonal change of variable also an orthogonal change of variable?

Exercise 8.113. Prove that under an orthogonal change of variable, the gradient in  $\vec{x}$  can be written in terms of the new variable by

$$\nabla f = \frac{\partial f}{\partial y_1} \frac{\vec{x}_{y_1}}{\|\vec{x}_{y_1}\|_2^2} + \frac{\partial f}{\partial y_2} \frac{\vec{x}_{y_2}}{\|\vec{x}_{y_2}\|_2^2} + \dots + \frac{\partial f}{\partial y_n} \frac{\vec{x}_{y_n}}{\|\vec{x}_{y_n}\|_2^2}.$$

### Elementary Symmetric Polynomial

The *elementary symmetric polynomials* for  $n$  variables  $x_1, x_2, \dots, x_n$  are

$$\sigma_k = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} x_{i_1} x_{i_2} \dots x_{i_k}, \quad k = 1, 2, \dots, n.$$

*Vieta's formulae* says that they appear as the coefficients of the polynomial

$$\begin{aligned} p(x) &= (x - x_1)(x - x_2) \dots (x - x_n) \\ &= x^n - \sigma_1 x^{n-1} + \sigma_2 x^{n-2} - \dots + (-1)^{n-1} \sigma_{n-1} x + (-1)^n \sigma_n. \end{aligned} \quad (8.7.1)$$

Therefore  $\vec{x} \mapsto \vec{\sigma}: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is the map that takes the roots of polynomials to polynomials.

Exercise 8.114. Prove that the derivative  $\frac{\partial \vec{\sigma}}{\partial \vec{x}}$  of the polynomial with respect to the roots satisfy

$$\begin{pmatrix} x_1^{n-1} & x_1^{n-2} & \dots & 1 \\ x_2^{n-1} & x_2^{n-2} & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{n-1} & x_n^{n-2} & \dots & 1 \end{pmatrix} \frac{\partial \vec{\sigma}}{\partial \vec{x}} + \begin{pmatrix} p'(x_1) & 0 & \dots & 0 \\ 0 & p'(x_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p'(x_n) \end{pmatrix} = O.$$

Then prove that when the roots  $\vec{x}_0$  of a polynomial  $p_0(x)$  are distinct, the roots  $\vec{x}$  of polynomials  $p(x)$  near  $p_0(x)$  is a continuously differentiable functions of the polynomial  $p(x)$ .

Exercise 8.115. Suppose  $x_0$  is a root of a polynomial  $p_0(x)$ . Prove that if  $x_0$  is not a multiple root (which is equivalent to  $p'_0(x_0) \neq 0$ ), then polynomials  $p(x)$  near  $p_0(x)$  have unique roots  $x(p)$  close to  $x_0$ . Moreover, the map  $p \mapsto x(p)$  is continuously differentiable.

### Power Sum and Newton's Identity

The *power sums* for  $n$  variables  $x_1, x_2, \dots, x_n$  are

$$s_k = \sum_{1 \leq i \leq n} x_i^k = x_1^k + x_2^k + \cdots + x_n^k, \quad k = 1, 2, \dots, n.$$

For the polynomial  $p(x)$  in (8.7.1), by adding  $p(x_i) = 0$  together for  $i = 1, 2, \dots, n$ , we get

$$s_n - \sigma_1 s_{n-1} + \sigma_2 s_{n-2} - \cdots + (-1)^{n-1} \sigma_{n-1} s_1 + (-1)^n n \sigma_n = 0. \quad (8.7.2)$$

Exercise 8.116. For

$$u_{l,k} = \sum_{\substack{i_1 < i_2 < \cdots < i_l \\ j \neq i_p}} x_{i_1} x_{i_2} \cdots x_{i_l} x_j^k, \quad l \geq 0, k \geq 1, l+k \leq n,$$

prove that

$$\begin{aligned} s_k &= u_{0,k}, \\ \sigma_1 s_{k-1} &= u_{0,k} + u_{1,k-1}, \\ \sigma_2 s_{k-2} &= u_{1,k-1} + u_{2,k-2}, \\ &\vdots \\ \sigma_{k-2} s_2 &= u_{k-3,3} + u_{k-2,2}, \\ \sigma_{k-1} s_1 &= u_{k-2,2} + k \sigma_k. \end{aligned}$$

and derive Newton's identities

$$s_k - \sigma_1 s_{k-1} + \sigma_2 s_{k-2} - \cdots + (-1)^{k-1} \sigma_{k-1} s_1 + (-1)^k k \sigma_k = 0, \quad k = 1, 2, \dots, n. \quad (8.7.3)$$

Exercise 8.117. Prove that there is a polynomial invertible map that relates  $\vec{s} = (s_1, s_2, \dots, s_n)$  and  $\vec{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_n)$ . Then discuss the local invertibility of the map  $\vec{x} \mapsto \vec{\sigma}: \mathbb{R}^n \rightarrow \mathbb{R}^n$  when there are multiple roots (see Exercise 8.114).

### Functional Dependence

A collection of functions are *functionally dependent* if some can be written as functions of the others. For example, the functions  $f = x + y$ ,  $g = x^2 + y^2$ ,  $h = x^3 + y^3$  are functionally dependent by  $h = \frac{3}{2}fg - \frac{1}{2}f^3$ . In the following exercises (except Exercise 8.118), all functions are continuously differentiable.

Exercise 8.118. At the purely set theoretical level, for two given maps  $f$  and  $g$  on  $X$ , what is the condition for the existence of a map  $h$  satisfying  $f(x) = h(g(x))$  for all  $x \in X$ ?

Exercise 8.119. Prove that if  $f_1(\vec{x}), f_2(\vec{x}), \dots, f_n(\vec{x})$  are functionally dependent, then the gradients  $\nabla f_1, \nabla f_2, \dots, \nabla f_n$  are linearly dependent.

Exercise 8.120. Prove that  $f_1(\vec{x}), f_2(\vec{x}), \dots, f_n(\vec{x})$  are functionally dependent near  $\vec{x}_0$  if and only if there is a function  $h(\vec{y})$  defined near  $\vec{y}_0 = (f_1(\vec{x}_0), f_2(\vec{x}_0), \dots, f_n(\vec{x}_0))$ , such that  $\nabla h(\vec{y}_0) \neq \vec{0}$  and  $h(f_1(\vec{x}), f_2(\vec{x}), \dots, f_n(\vec{x})) = 0$ .

Exercise 8.121. Suppose  $\nabla g(\vec{x}_0) \neq \vec{0}$ , and  $\nabla f$  is always a scalar multiple of  $\nabla g$  near  $\vec{x}_0$ . Prove that there is a function  $h(y)$  defined for  $y$  near  $g(\vec{x}_0)$ , such that  $f(\vec{x}) = h(g(\vec{x}))$  near  $\vec{x}_0$ .

Hint: If  $\frac{\partial g}{\partial x_1} \neq 0$ , then  $(x_1, x_2, \dots, x_n) \mapsto (g, x_2, \dots, x_n)$  is invertible near  $\vec{x}_0$ . After changing the variables from  $(x_1, x_2, \dots, x_n)$  to  $(g, x_2, \dots, x_n)$ , verify that  $\frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0$ .

Exercise 8.122. Suppose  $\nabla g_1, \nabla g_2, \dots, \nabla g_k$  are linearly independent at  $\vec{x}_0$ , and  $\nabla f$  is a linear combination of  $\nabla g_1, \nabla g_2, \dots, \nabla g_k$  near  $\vec{x}_0$ . Prove that there is a function  $h(\vec{y})$  defined for  $\vec{y}$  near  $(g_1(\vec{x}_0), g_2(\vec{x}_0), \dots, g_k(\vec{x}_0))$ , such that  $f(\vec{x}) = h(g_1(\vec{x}), g_2(\vec{x}), \dots, g_k(\vec{x}))$  near  $\vec{x}_0$ .

Exercise 8.123. Suppose the rank of the gradient vectors  $\nabla f_1, \nabla f_2, \dots, \nabla f_m$  is always  $k$  near  $\vec{x}_0$ . Prove that there are  $k$  functions from  $f_1, f_2, \dots, f_m$ , such that the other  $m - k$  functions are functionally dependent on these  $k$  functions.

Exercise 8.124. Determine functional dependence.

1.  $x + y + z, x^2 + y^2 + z^2, x^3 + y^3 + z^3$ .
2.  $x + y - z, x - y + z, x^2 + y^2 + z^2 - 2yz$ .
3.  $\frac{x}{x^2 + y^2 + z^2}, \frac{y}{x^2 + y^2 + z^2}, \frac{z}{x^2 + y^2 + z^2}$ .
4.  $\frac{x}{\sqrt{x^2 + y^2 + z^2}}, \frac{y}{\sqrt{x^2 + y^2 + z^2}}, \frac{z}{\sqrt{x^2 + y^2 + z^2}}$ .

### Convex Subset and Convex Function

A subset  $A \subset \mathbb{R}^n$  is convex if  $\vec{x}, \vec{y} \in A$  implies the straight line connecting  $\vec{x}$  and  $\vec{y}$  still lies in  $A$ . In other words,  $(1 - \lambda)\vec{x} + \lambda\vec{y} \in A$  for any  $0 < \lambda < 1$ .

A function  $f(\vec{x})$  defined on a convex subset  $A$  is convex if

$$0 < \lambda < 1 \implies (1 - \lambda)f(\vec{x}) + \lambda f(\vec{y}) \geq f((1 - \lambda)\vec{x} + \lambda\vec{y}).$$

Exercise 8.125. For any  $\lambda_1, \lambda_2, \dots, \lambda_k$  satisfying  $\lambda_1 + \lambda_2 + \dots + \lambda_k = 1$  and  $0 \leq \lambda_i \leq 1$ , extend *Jensen's inequality* in Exercise 3.118 to multivariable convex functions

$$f(\lambda_1 \vec{x}_1 + \lambda_2 \vec{x}_2 + \dots + \lambda_k \vec{x}_k) \leq \lambda_1 f(\vec{x}_1) + \lambda_2 f(\vec{x}_2) + \dots + \lambda_k f(\vec{x}_k).$$

Exercise 8.126. Prove that a function  $f(\vec{x})$  is convex if and only if its restriction on any straight line is convex.

Exercise 8.127. Prove that a function  $f(\vec{x})$  is convex if and only if for any linear function  $L(\vec{x}) = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$ , the subset  $\{\vec{x}: f(\vec{x}) \leq L(\vec{x})\}$  is convex.

Exercise 8.128. Extend Exercise 3.112 to multivariable: A function  $f(\vec{x})$  defined on an open convex subset is convex if and only if for any  $\vec{z}$  in the subset, there is a linear function  $K(\vec{x})$ , such that  $K(\vec{z}) = f(\vec{z})$  and  $K(\vec{x}) \leq f(\vec{x})$ .

Exercise 8.129. Prove that any convex function on an open convex subset is continuous.

Exercise 8.130. Prove that a second order continuously differentiable function  $f(\vec{x})$  on an open convex subset is convex if and only if the Hessian is positive semi-definite:  $h_f(\vec{v}) \geq 0$  for any  $\vec{v}$ .

Remark on Alexandrov theorem ????????

### Laplacian

The *Laplacian* of a function  $f(\vec{x})$  is

$$\Delta f = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}.$$

Functions satisfying the *Laplace equation*  $\Delta f = 0$  are called *harmonic*.

Exercise 8.131. Prove that  $\Delta(f + g) = \Delta f + \Delta g$  and  $\Delta(fg) = g\Delta f + f\Delta g + 2\nabla f \cdot \nabla g$ .

Exercise 8.132. Suppose a function  $f(\vec{x}) = u(r)$  depends only on the Euclidean norm  $r = \|\vec{x}\|_2$  of the vector. Prove that  $\Delta f = u''(r) + (n-1)r^{-1}u'(r)$  and find the condition for the function to be harmonic.

Exercise 8.133. Derive the Laplacian in  $\mathbb{R}^2$

$$\Delta f = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r} \frac{\partial f}{\partial r} + \frac{1}{r^2} \frac{\partial^2 f}{\partial \theta^2}.$$

in the polar coordinates. Also derive the Laplacian in  $\mathbb{R}^3$

$$\Delta f = \frac{\partial^2 f}{\partial r^2} + \frac{2}{r} \frac{\partial f}{\partial r} + \frac{1}{r^2} \frac{\partial^2 f}{\partial \theta^2} + \frac{\cos \phi}{r^2 \sin \phi} \frac{\partial f}{\partial \theta} + \frac{1}{r^2 \sin^2 \phi} \frac{\partial^2 f}{\partial \theta^2}.$$

in the spherical coordinates.

Exercise 8.134. Let  $\vec{x} = F(\vec{y}): \mathbb{R}^n \rightarrow \mathbb{R}^n$  be an orthogonal change of variable (see Exercise (8.112)). Let  $|J| = \|\vec{x}_{y_1}\|_2 \|\vec{x}_{y_2}\|_2 \cdots \|\vec{x}_{y_n}\|_2$  be the absolute value of the determinant of the Jacobian matrix. Prove the Laplacian is

$$\Delta f = \frac{1}{|J|} \left( \frac{\partial}{\partial y_1} \left( \frac{|J|}{\|\vec{x}_{y_1}\|_2^2} \frac{\partial f}{\partial y_1} \right) + \frac{\partial}{\partial y_2} \left( \frac{|J|}{\|\vec{x}_{y_2}\|_2^2} \frac{\partial f}{\partial y_2} \right) + \cdots + \frac{\partial}{\partial y_n} \left( \frac{|J|}{\|\vec{x}_{y_n}\|_2^2} \frac{\partial f}{\partial y_n} \right) \right).$$

In particular, if the change of variable is orthonormal (i.e.,  $\|\vec{x}_{y_i}\|_2 = 1$ ), then the Laplacian is not changed.

### Euler Equation

Exercise 8.135. Suppose a function  $f$  is differentiable away from  $\vec{0}$ . Prove that  $f$  is homogeneous of degree  $p$  if and only if it satisfies the *Euler equation*

$$x_1 f_{x_1} + x_2 f_{x_2} + \cdots + x_n f_{x_n} = pf.$$

What about a weighted homogeneous function?

Exercise 8.136. Extend the Euler equation to high order derivatives

$$\sum_{i_1, i_2, \dots, i_k \geq 1} x_{i_1} x_{i_2} \cdots x_{i_k} \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} = p(p-1) \cdots (p-k+1)f.$$

Exercise 8.137. Prove that a change of variable  $\vec{x} = H(\vec{z}): \mathbb{R}^n \rightarrow \mathbb{R}^n$  preserves  $\sum x_i f_{x_i}$

$$z_1(f \circ H)_{z_1} + z_2(f \circ H)_{z_2} + \cdots + z_n(f \circ H)_{z_n} = x_1 f_{x_1} + x_2 f_{x_2} + \cdots + x_n f_{x_n}.$$

for any differentiable function  $f(\vec{x})$  if and only if it preserves the scaling

$$H(c\vec{z}) = cH(\vec{z}) \text{ for any } c > 0.$$

### Size of Polynomial vs. Size of Root

In Exercises 8.114 and 8.115, polynomials are identified with the coefficient vector  $\vec{\sigma}$ . The roots can often be locally considered as functions of the polynomial. The following exercises explore the relation between the size of  $\vec{\sigma}$  (which is the size of the polynomial) and the size of the root  $x$ .

Exercise 8.138. Prove that  $x(\vec{\sigma})$  is weighted homogeneous

$$tx(\sigma_1, \sigma_2, \dots, \sigma_n) = x(t\sigma_1, t^2\sigma_2, \dots, t^n\sigma_n).$$

Then use this to show that the root function has no local extreme for  $\vec{\sigma} \in \mathbb{R}^n$ .

Exercise 8.139. Among all the polynomials with coefficients satisfying  $\|\vec{\sigma}\|_2 \leq 1$ , what is the largest root? Moreover, does the root have any local maximum?

## **Chapter 9**

# **Measure**

## 9.1 Length in $\mathbb{R}$

Starting from the length of an intervals, we try to extend the length to more good subsets of  $\mathbb{R}$ .

### Length of Finite Union of Intervals

Denote by  $\langle a, b \rangle$  an interval of left end  $a$  and right end  $b$ . The notation can mean any one of  $(a, b)$ ,  $[a, b]$ ,  $(a, b]$  or  $[a, b)$ . The *length* of the interval is

$$\lambda\langle a, b \rangle = b - a.$$

More generally, for a (pairwise) disjoint union

$$\sqcup_{i=1}^n \langle a_i, b_i \rangle = \langle a_1, b_1 \rangle \sqcup \langle a_2, b_2 \rangle \cdots \sqcup \langle a_n, b_n \rangle$$

of finitely many intervals, the length is the sum of the lengths of the intervals

$$\lambda(\sqcup_{i=1}^n \langle a_i, b_i \rangle) = \sum_{i=1}^n (b_i - a_i).$$

In the future, “disjoint” will always mean “pairwise disjoint”. A union denoted by  $\sqcup$  will always mean (pairwise) disjoint union.

**Proposition 9.1.1.** *If*

$$\langle a_1, b_1 \rangle \sqcup \langle a_2, b_2 \rangle \sqcup \cdots \sqcup \langle a_m, b_m \rangle \subset \langle c_1, d_1 \rangle \cup \langle c_2, d_2 \rangle \cup \cdots \cup \langle c_n, d_n \rangle,$$

*then*

$$\lambda(\sqcup_{i=1}^m \langle a_i, b_i \rangle) = \sum_{i=1}^m \lambda\langle a_i, b_i \rangle \leq \sum_{j=1}^n \lambda\langle c_j, d_j \rangle.$$

Since there might be some overlapping between different intervals  $\langle c_j, d_j \rangle$ , the sum  $\sum_{j=1}^n \lambda\langle c_j, d_j \rangle$  on the right side is not necessarily the length of the union  $\langle c_1, d_1 \rangle \cup \langle c_2, d_2 \rangle \cup \cdots \cup \langle c_n, d_n \rangle$ .

As an immediate consequence of the proposition, if  $\sqcup_{i=1}^m \langle a_i, b_i \rangle = \sqcup_{j=1}^n \langle c_j, d_j \rangle$ , then the equality means  $\subset$  and  $\supset$ . By applying the proposition to  $\subset$  and to  $\supset$ , we get  $\sum_{i=1}^m (b_i - a_i) = \sum_{j=1}^n (d_j - c_j)$ . This shows that the definition of  $\lambda(\sqcup_{i=1}^m \langle a_i, b_i \rangle)$  is not ambiguous.

*Proof.* If  $k = 1$ , then

$$\langle a_1, b_1 \rangle \sqcup \langle a_2, b_2 \rangle \sqcup \cdots \sqcup \langle a_m, b_m \rangle \subset \langle c_1, d_1 \rangle.$$

The disjoint property implies that, after rearranging the intervals, we may assume

$$c_1 \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq \cdots \leq a_m \leq b_m \leq d_1.$$



Then

$$\begin{aligned}\lambda(\sqcup_{i=1}^m \langle a_i, b_i \rangle) &= \sum_{i=1}^m (b_i - a_i) \\ &= b_m - (a_m - b_{m-1}) - \cdots - (a_3 - b_2) - (a_2 - b_1) - a_1 \\ &\leq b_m - a_1 \leq d_1 - c_1 = \lambda\langle c_1, d_1 \rangle.\end{aligned}$$

Next assume the proposition holds for  $n-1$ . We try to prove the proposition for  $n$ . Note that  $c_n, d_n$  divide the whole real line into three disjoint parts,

$$\mathbb{R} = (-\infty, c_n) \sqcup \langle c_n, d_n \rangle \sqcup \langle d_n, +\infty \rangle,$$

and any interval is divided accordingly into three disjoint intervals,

$$\langle a, b \rangle = \langle a^-, b^- \rangle \sqcup \langle a', b' \rangle \sqcup \langle a^+, b^+ \rangle.$$

It is easy to verify that

$$\lambda\langle a, b \rangle = \lambda\langle a^-, b^- \rangle + \lambda\langle a', b' \rangle + \lambda\langle a^+, b^+ \rangle.$$

The inclusion  $\langle a_1, b_1 \rangle \sqcup \langle a_2, b_2 \rangle \cdots \sqcup \langle a_m, b_m \rangle \subset \langle c_1, d_1 \rangle \cup \langle c_2, d_2 \rangle \cup \cdots \cup \langle c_n, d_n \rangle$  implies that

$$\langle a'_1, b'_1 \rangle \sqcup \langle a'_2, b'_2 \rangle \cdots \sqcup \langle a'_m, b'_m \rangle \subset \langle c_n, d_n \rangle,$$

and

$$\begin{aligned}&\langle a_1^-, b_1^- \rangle \sqcup \langle a_2^-, b_2^- \rangle \cdots \sqcup \langle a_m^-, b_m^- \rangle \sqcup \langle a_1^+, b_1^+ \rangle \sqcup \langle a_2^+, b_2^+ \rangle \cdots \sqcup \langle a_m^+, b_m^+ \rangle \\ &\subset \langle c_1, d_1 \rangle \cup \langle c_2, d_2 \rangle \cup \cdots \cup \langle c_{n-1}, d_{n-1} \rangle.\end{aligned}$$

By the inductive assumption, we have

$$\sum_{i=1}^m [\lambda\langle a_i^-, b_i^- \rangle + \lambda\langle a_i^+, b_i^+ \rangle] \leq \sum_{j=1}^{n-1} \lambda\langle c_j, d_j \rangle,$$

and

$$\sum_{i=1}^m \lambda\langle a'_i, b'_i \rangle \leq \lambda\langle c_n, d_n \rangle.$$

Adding the inequalities together, we get

$$\begin{aligned}\lambda(\sqcup_{i=1}^m \langle a_i, b_i \rangle) &= \sum_{i=1}^m \lambda\langle a_i, b_i \rangle \\ &= \sum_{i=1}^m [\lambda\langle a_i^-, b_i^- \rangle + \lambda\langle a'_i, b'_i \rangle + \lambda\langle a_i^+, b_i^+ \rangle] \\ &\leq \sum_{j=1}^{n-1} \lambda\langle c_j, d_j \rangle + \lambda\langle c_n, d_n \rangle = \sum_{j=1}^n \lambda\langle c_j, d_j \rangle.\end{aligned}$$

This completes the proof by induction.  $\square$

It is easy to see that any union of finitely many intervals can be expressed as a (*pairwise*) *disjoint* union of finitely many intervals. Moreover, if  $A$  and  $B$  are disjoint unions of finitely many intervals, then  $A \cup B$ ,  $A \cap B$  and  $A - B$  are also disjoint unions of finitely many intervals. See Exercise 9.2 for details.

**Proposition 9.1.2.** *The length of disjoint unions of finitely many intervals satisfy*

$$\lambda(A \cup B) = \lambda(A) + \lambda(B) - \lambda(A \cap B).$$

*Proof.* The end points of intervals in  $A$  and  $B$  divide the real line into finitely many disjoint intervals. Then  $A$ ,  $B$ ,  $A \cup B$ ,  $A \cap B$  are unions of some of these disjoint intervals. More specifically, let  $\mathcal{I}_A$  and  $\mathcal{I}_B$  be the collections of such intervals included in  $A$  and  $B$ . Then

$$A = \sqcup_{I \in \mathcal{I}_A} I, \quad B = \sqcup_{I \in \mathcal{I}_B} I, \quad A \cup B = \sqcup_{I \in \mathcal{I}_A \cup \mathcal{I}_B} I, \quad A \cap B = \sqcup_{I \in \mathcal{I}_A \cap \mathcal{I}_B} I.$$

By the definition of the length of union of disjoint intervals,

$$\begin{aligned} \lambda(A \cup B) &= \sum_{I \in \mathcal{I}_A \cup \mathcal{I}_B} \lambda(I) = \sum_{I \in \mathcal{I}_A} \lambda(I) + \sum_{I \in \mathcal{I}_B} \lambda(I) - \sum_{I \in \mathcal{I}_A \cap \mathcal{I}_B} \lambda(I) \\ &= \lambda(A) + \lambda(B) - \lambda(A \cap B). \end{aligned}$$

Strictly speaking, the first and the third equalities use the fact that the concept of length is well defined. The reason for the second equality is that the sum  $\sum_{I \in \mathcal{I}_A} \lambda(I) + \sum_{I \in \mathcal{I}_B} \lambda(I)$  double counts the lengths in the sum  $\sum_{I \in \mathcal{I}_A \cup \mathcal{I}_B} \lambda(I)$  of those intervals in  $\mathcal{I}_A \cap \mathcal{I}_B$ . Therefore subtracting the sum  $\sum_{I \in \mathcal{I}_A \cap \mathcal{I}_B} \lambda(I)$  of double counted intervals restores the equality.  $\square$

**Exercise 9.1.** Let  $\Sigma$  be the collection of disjoint unions of finitely many intervals. The following steps establish the property that  $A, B \in \Sigma$  implies  $A \cup B, A \cap B, A - B \in \Sigma$ .

1. Prove that if  $A, B \in \Sigma$ , then the intersection  $A \cap B \in \Sigma$ .
2. Prove that if  $A \in \Sigma$ , then the complement  $\mathbb{R} - A \in \Sigma$ .
3. Use the first two steps to prove that if  $A, B \in \Sigma$ , then the union  $A \cup B \in \Sigma$  and the subtraction  $A - B \in \Sigma$ .

**Exercise 9.2.** Let  $\mathcal{C}$  be a collection of subsets of  $X$ . Let  $\Sigma$  be the collection of disjoint unions of finitely many subsets from  $\mathcal{C}$ . Prove that if  $A, B \in \mathcal{C}$  implies  $A \cap B, X - A \in \Sigma$ , then  $A, B \in \Sigma$  implies  $A \cup B, A \cap B, A - B \in \Sigma$ .

**Exercise 9.3.** Prove that if  $A, B, C$  are disjoint unions of finitely many intervals, then

$$\begin{aligned} \lambda(A \cup B \cup C) &= \lambda(A) + \lambda(B) + \lambda(C) \\ &\quad - \lambda(A \cap B) - \lambda(B \cap C) - \lambda(C \cap A) + \lambda(A \cap B \cap C). \end{aligned}$$

**Exercise 9.4.** Let  $A_i$  be disjoint unions of finitely many intervals. Prove  $\lambda(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \lambda(A_i)$ , and that the inequality becomes equality if  $A_i$  are disjoint.

## Length of Open Subsets

A subset  $U \subset \mathbb{R}$  is *open* if  $x \in U$  implies  $(x - \epsilon, x + \epsilon) \subset U$  for some  $\epsilon > 0$ . See Definition 6.4.1. By Theorem 6.4.6,  $U$  is a disjoint union of *countably* many open intervals, and the decomposition is unique.

**Definition 9.1.3.** The *length* of a bounded open subset  $U = \sqcup(a_i, b_i) \subset \mathbb{R}$  is

$$\lambda(U) = \sum (b_i - a_i).$$

The bounded open subset  $U$  is contained in a bounded interval  $[c, d]$ . Then we have  $\sqcup_{i=1}^n (a_i, b_i) \subset [c, d]$  for any  $n$ , and Proposition 9.1.1 implies that the partial sums of the series  $\sum (b_i - a_i)$  are bounded. Therefore the series in the definition converges. Moreover, the definition clearly extends the length of unions of finitely many open intervals.

By Proposition 6.4.4, unions of open subsets are open, and finite intersections of open subsets are also open. The following extends Propositions 9.1.1 and 9.1.2.

**Proposition 9.1.4.** *The length of open subsets has the following properties.*

1. If  $U \subset \cup V_i$ , then  $\lambda(U) \leq \sum \lambda(V_i)$ .
2.  $\lambda(U \cup V) = \lambda(U) + \lambda(V) - \lambda(U \cap V)$ .

The first property implies that  $\lambda$  is *monotone*

$$U \subset V \implies \lambda(U) \leq \lambda(V),$$

and *countably subadditive*

$$\lambda(\cup U_i) \leq \sum \lambda(U_i).$$

*Proof.* By expressing each  $V_i$  as a disjoint union of open intervals, the first property means that  $U \subset \cup (c_j, d_j)$  implies  $\lambda(U) \leq \sum (d_j - c_j)$ . Let  $U = \sqcup(a_i, b_i)$ . Then for any  $\epsilon > 0$  and  $n$ , the compact subset

$$K = [a_1 + \epsilon, b_1 - \epsilon] \sqcup \cdots \sqcup [a_n + \epsilon, b_n - \epsilon]$$

is covered by the open intervals  $(c_j, d_j)$ . By Theorem 1.5.6, we have

$$K \subset (c_1, d_1) \cup \cdots \cup (c_k, d_k)$$

for finitely many intervals among  $(c_j, d_j)$ . Applying Proposition 9.1.1, we get

$$\sum_{i=1}^n (b_i - a_i) - 2n\epsilon = \sum_{i=1}^n [(b_i - \epsilon) - (a_i + \epsilon)] \leq \sum_{j=1}^k (d_j - c_j) \leq \sum (d_j - c_j).$$

For fixed  $n$ , since  $\epsilon$  is arbitrary and independent of  $n$ , we get

$$\sum_{i=1}^n (b_i - a_i) \leq \sum (d_j - c_j).$$

Then since the right side is independent of  $n$ , we conclude that

$$\lambda(U) = \sum (b_i - a_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n (b_i - a_i) \leq \sum (d_j - c_j).$$

For the second property, let  $U = \sqcup (a_i, b_i)$ ,  $V = \sqcup (c_j, d_j)$ , and

$$\begin{aligned} U_n &= (a_1, b_1) \sqcup \cdots \sqcup (a_n, b_n), \\ V_n &= (c_1, d_1) \sqcup \cdots \sqcup (c_n, d_n). \end{aligned}$$

We note that  $U - U_n = \sqcup_{i>n} (a_i, b_i)$  is still open, and  $V - V_n$  is also open. By Proposition 9.1.2, we have (recall that  $\lambda$  extends  $l$ )

$$\lambda(U_n \cup V_n) = \lambda(U_n) + \lambda(V_n) - \lambda(U_n \cap V_n).$$

We will argue about the limits of the four terms in the equality. Then taking the limit as  $n \rightarrow \infty$  gives us

$$\lambda(U \cup V) = \lambda(U) + \lambda(V) - \lambda(U \cap V).$$

By the convergence of  $\sum (b_i - a_i)$  and  $\sum (d_i - c_i)$ , for any  $\epsilon > 0$ , there is  $N$ , such that  $n > N$  implies

$$\begin{aligned} 0 \leq \lambda(U) - \lambda(U_n) &= \lambda(U - U_n) = \sum_{i>n} (b_i - a_i) \leq \epsilon, \\ 0 \leq \lambda(V) - \lambda(V_n) &= \lambda(V - V_n) = \sum_{i>n} (d_i - c_i) \leq \epsilon. \end{aligned}$$

Moreover, by

$$\begin{aligned} U_n \cup V_n &\subset U \cup V = (U_n \cup V_n) \cup (U - U_n) \cup (V - V_n), \\ U_n \cap V_n &\subset U \cap V \subset (U_n \cap V_n) \cup (U - U_n) \cup (V - V_n), \end{aligned}$$

and the first part,  $n > N$  also implies

$$\begin{aligned} \lambda(U_n \cup V_n) &\leq \lambda(U \cup V) \\ &\leq \lambda(U_n \cup V_n) + \lambda(U - U_n) + \lambda(V - V_n) \\ &\leq \lambda(U_n \cup V_n) + 2\epsilon, \\ \lambda(U_n \cap V_n) &\leq \lambda(U \cap V) \\ &\leq \lambda(U_n \cap V_n) + \lambda(U - U_n) + \lambda(V - V_n) \\ &\leq \lambda(U_n \cap V_n) + 2\epsilon. \end{aligned}$$

Therefore we conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} \lambda(U_n) &= \lambda(U), & \lim_{n \rightarrow \infty} \lambda(V_n) &= \lambda(V), \\ \lim_{n \rightarrow \infty} \lambda(U_n \cup V_n) &= \lambda(U \cup V), & \lim_{n \rightarrow \infty} \lambda(U_n \cap V_n) &= \lambda(U \cap V). \end{aligned} \quad \square$$

## Length of Compact Subsets

We will use the length of bounded open subsets to approximate any bounded subsets of  $\mathbb{R}$ . Specifically, we consider all open subsets  $U$  containing a bounded subset  $A$ . The “length” of  $A$ , whatever our future definition is, should be no more than  $\lambda(U)$ . Therefore, we get the upper bound

$$\mu^*(A) = \inf\{\lambda(U) : A \subset U, U \text{ open}\}$$

for the length of  $A$ .

However, in such consideration, we also note that  $A \subset U \subset [-r, r]$  is the same as  $K = [-r, r] - U \subset [-r, r] - A = B$ , where  $B$  can also be any bounded subset, and  $K$  is closed. This means that, as a consequence of using open subsets to approximate subsets *from outside*, we should also use closed subsets to approximate subsets *from inside*. Such consideration makes it necessary to define the length of closed subsets.

**Definition 9.1.5.** The *length* of a bounded and closed (same as compact) subset  $K \subset \mathbb{R}$  is

$$\lambda(K) = \lambda(U) - \lambda(U - K),$$

where  $U$  is a bounded open subset containing  $K$ .

We need to verify that the definition is independent of the choice of  $U$ . Let  $U$  and  $V$  be open subsets containing  $K$ , then  $W = U \cap V$  is an open subset satisfying  $K \subset W \subset U$ . Applying the second property in Proposition 9.1.4 to

$$U = (U - K) \cup W, \quad W - K = (U - K) \cap W,$$

we get

$$\lambda(U) = \lambda(U - K) + \lambda(W) - \lambda(W - K).$$

This is the same as

$$\lambda(U) - \lambda(U - K) = \lambda(W) - \lambda(W - K).$$

By the same reason, we also have

$$\lambda(V) - \lambda(V - K) = \lambda(W) - \lambda(W - K).$$

This proves that the definition of  $\lambda(K)$  is independent of  $U$ .

**Exercise 9.5.** Prove that the definition of the length of bounded and closed subsets extends the length of unions of finitely many closed intervals.

**Exercise 9.6.** Prove properties of the length of bounded and closed subsets.

1.  $K \subset L$  implies  $\lambda(K) \leq \lambda(L)$ .
2.  $\lambda(K \cup L) = \lambda(K) + \lambda(L) - \lambda(K \cap L)$ .

**Exercise 9.7.** Suppose  $K$  is compact and  $U$  is open.

1. Prove that  $\lambda(K) \leq \lambda(K - U) + \lambda(U)$ , and equality happens when  $U \subset K$ .
2. Prove that  $\lambda(U) \leq \lambda(U - K) + \lambda(K)$ , and the equality happens when  $K \subset U$ .

**Exercise 9.8.** Prove that if  $U \subset K \subset V$ ,  $K$  compact,  $U$  and  $V$  open, then  $\lambda(U) \leq \lambda(K) \leq \lambda(V)$ .

## 9.2 Lebesgue Measure in $\mathbb{R}$

The approximation of any bounded subset from outside by open subsets and from inside by closed subsets leads to the upper and lower bounds of the length of the subset. When the two bounds coincide, there is no ambiguity about the length of the subset.

**Definition 9.2.1.** The *Lebesgue outer measure* of a bounded subset  $A \subset \mathbb{R}$  is

$$\mu^*(A) = \inf\{\lambda(U) : A \subset U, U \text{ open}\}.$$

The *Lebesgue inner measure* is

$$\mu_*(A) = \sup\{\lambda(K) : K \subset A, K \text{ closed}\}.$$

The subset is *Lebesgue measurable* if  $\mu^*(A) = \mu_*(A)$  and the number is the *Lebesgue measure*  $\mu(A)$  of  $A$ .

It follows immediately that  $A$  is Lebesgue measurable if and only if for any  $\epsilon > 0$ , there are open  $U$  and closed  $K$ , such that

$$K \subset A \subset U \text{ and } \lambda(U - K) = \lambda(U) - \lambda(K) < \epsilon.$$

Moreover, the measure  $\mu(A)$  is the only number satisfying  $\lambda(K) \leq \mu(A) \leq \lambda(U)$  for any  $K \subset A \subset U$ .

**Proposition 9.2.2.** *The outer and inner measures have the following properties.*

1.  $0 \leq \mu_*(A) \leq \mu^*(A)$ .
2.  $A \subset B$  implies  $\mu^*(A) \leq \mu^*(B)$ ,  $\mu_*(A) \leq \mu_*(B)$ .
3.  $\mu^*(\cup A_i) \leq \sum \mu^*(A_i)$  for any countable union  $\cup A_i$ .

*Proof.* The inequality  $\mu_*(A) \leq \mu^*(A)$  follows from  $\lambda(K) = \lambda(U) - \lambda(U - K) \leq \lambda(U)$  for  $K \subset A \subset U$ . The second property follows from the definition and the monotone property of  $\lambda$  (Proposition 9.1.4 and Exercise 9.6).

Now we prove the third property. For any  $\epsilon > 0$  and  $i$ , by the definition of  $\mu^*$ , there is open  $U_i$ , such that

$$A_i \subset U_i, \quad \lambda(U_i) < \mu^*(A_i) + \frac{\epsilon}{2^i}.$$

Then  $\cup A_i \subset \cup U_i$ , the union  $\cup U_i$  of open subsets is still open, and by the countable subadditivity of  $\lambda$ , we have

$$\mu^*(\cup A_i) \leq \lambda(\cup U_i) \leq \sum \lambda(U_i) < \sum \left( \mu^*(A_i) + \frac{\epsilon}{2^i} \right) \leq \sum \mu^*(A_i) + \epsilon.$$

Since  $\epsilon$  is arbitrary, we get  $\mu^*(\cup A_i) \leq \sum \mu^*(A_i)$ .  $\square$

An immediate consequence of the proposition is the following result on subsets of measure 0.

**Proposition 9.2.3.** *If  $\mu^*(A) = 0$ , then  $A$  is measurable with  $\mu(A) = 0$ . Moreover, any subset of a set of measure zero is also a set of measure zero.*

Exercises 9.12 through 9.18 establish basic properties of the Lebesgue measure in  $\mathbb{R}$ . These properties will be established again in more general context in the subsequent sections.

**Example 9.2.1.** For any interval  $A = \langle a, b \rangle$ , the inclusions

$$K = [a + \epsilon, b - \epsilon] \subset A \subset U = (a - \epsilon, b + \epsilon)$$

tell us

$$(b - a) - 2\epsilon \leq \mu_*(A) \leq \mu^*(A) \leq (b - a) + 2\epsilon.$$

Since  $\epsilon$  is arbitrary, we get  $\mu_*(A) = \mu^*(A) = b - a$ . Therefore the interval is Lebesgue measurable, with the usual length as the measure.

**Example 9.2.2.** Let  $A = \{a_n : n \in \mathbb{N}\}$  be a countable subset. Then

$$\mu^*(A) = \mu^*(\cup \{a_n\}) \leq \sum \mu^*(\{a_n\}) = \sum 0 = 0.$$

Therefore any countable subset has measure 0. In particular, this implies that the interval  $[0, 1]$  is not countable.

**Example 9.2.3.** We should verify that the general definition of Lebesgue measure extends the length of open subsets defined earlier. For any open subset  $A = \sqcup (a_i, b_i)$ , the inclusions

$$K = [a_1 + \epsilon, b_1 - \epsilon] \cup \cdots \cup [a_n + \epsilon, b_n - \epsilon] \subset A \subset U = A$$

tell us

$$\sum_{i=1}^n (b_i - a_i) - 2n\epsilon \leq \mu_*(A) \leq \mu^*(A) \leq \sum_{i=1}^{\infty} (b_i - a_i).$$

Since this is true for any  $\epsilon$  and  $n$ , we get

$$\mu_*(A) = \mu^*(A) = \sum (b_i - a_i) = \lambda(A).$$

The Lebesgue measure also extends the length of closed subsets. See Exercise 9.13.

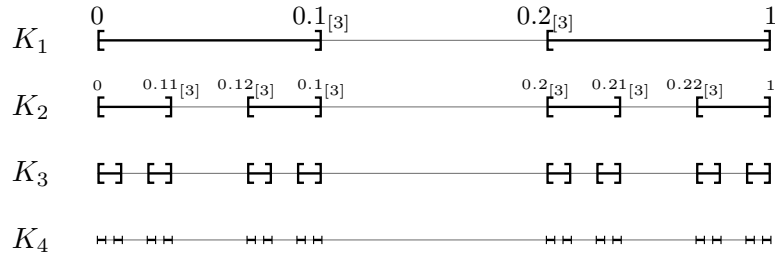
**Example 9.2.4 (Cantor Set).** Example 9.2.2 tells us that countable subsets have measure zero. Here we construct an uncountable subset with measure zero.

For any closed interval  $[a, b]$ , we delete the middle third segment to get

$$[a, b]^* = [a, b] - \left( \frac{a+2b}{3}, \frac{2a+b}{3} \right) = \left[ a, \frac{a+2b}{3} \right] \cup \left[ \frac{2a+b}{3}, b \right].$$

For a finite union  $A = \sqcup [a_i, b_i]$  of closed intervals, denote  $A^* = \sqcup [a_i, b_i]^*$ . Define  $K_0 = [0, 1]$ ,  $K_{n+1} = K_n^*$ . Then the *Cantor set* is

$$K = \cap K_n = [0, 1] - \left( \frac{1}{3}, \frac{2}{3} \right) - \left( \frac{1}{9}, \frac{2}{9} \right) - \left( \frac{7}{9}, \frac{8}{9} \right) - \left( \frac{1}{27}, \frac{2}{27} \right) - \cdots.$$



**Figure 9.2.1.** construction of Cantor set

The Cantor set  $K$  is closed because it is the complement of an open subset. Since  $K$  is obtained by deleting 1 interval of length  $\frac{1}{3}$ , 2 interval of length  $\frac{1}{3^2}$ ,  $2^2 = 4$  interval of length  $\frac{1}{3^3}$ ,  $\dots$ ,  $2^{n-1}$  interval of length  $\frac{1}{3^n}$ ,  $\dots$ , by first part of Exercise 9.7, we have

$$\lambda(K) = \lambda[0, 1] - \left( \frac{1}{3} + 2 \frac{1}{3^2} + 2^2 \frac{1}{3^3} + \cdots \right) = 1 - \frac{1}{3} \frac{1}{1 - \frac{1}{3}} = 0.$$

Exercise 9.10 gives another argument for  $\lambda(K) = 0$  that does not use Exercise 9.7.

The middle third of the interval  $[0, 1]$  consists of numbers of the form  $0.1 \cdots$  in base 3. Therefore deleting the interval means numbers of the form  $0.0 \cdots_{[3]}$  or  $0.2 \cdots_{[3]}$ . (Note that  $\frac{1}{3} = 0.0\bar{2}_{[3]}$  and  $1 = 0.\bar{2}_{[3]}$ , where  $\bar{2}$  means 2 repeated forever.) The same pattern repeats for  $\left[0, \frac{1}{3}\right]$ , which consists of numbers of the form  $0.0 \cdots_{[3]}$ . Deleting the middle third from  $\left[0, \frac{1}{3}\right]$  gives us the numbers of the form  $0.00 \cdots_{[3]}$  or  $0.02 \cdots_{[3]}$ . Similarly, deleting the middle third from  $\left[\frac{2}{3}, 1\right]$  gives us  $0.20 \cdots_{[3]}$  or  $0.22 \cdots_{[3]}$ . The pattern repeats, and we end up with the base 3 description of the Cantor set

$$K = \{0.a_1a_2a_3 \cdots_{[3]} : a_i = 0 \text{ or } 2\}, \quad 0.a_1a_2a_3 \cdots_{[3]} = \sum_{i=1}^{\infty} \frac{a_i}{3^i}.$$

By expressing numbers in  $[0, 1]$  in binary form (i.e., in base 2), we get a map

$$\kappa: K \rightarrow [0, 1], \quad \kappa(0.a_1a_2a_3 \cdots_{[3]}) = 0.b_1b_2b_3 \cdots_{[2]}, \quad a_i = 0 \text{ or } 2, \quad b_i = \frac{a_i}{2} = 0 \text{ or } 1.$$



The map is generally one-to-one, with the only exception that the two ends of any interval in  $[0, 1] - K$  should give the same value (and thus all exceptions are two-to-one)

$$\kappa(0.a_1 \cdots a_n 0\bar{2}_{[3]}) = \kappa(0.a_1 \cdots a_n 2_{[3]}) = 0.b_1 \cdots b_n 1_{[2]}.$$

In particular, the number of elements in  $K$  is the same as the real numbers in  $[0, 1]$ . This number is not countable.

**Exercise 9.9.** Directly prove that any countable subset has measure 0. In other words, for any countable subset  $A$  and  $\epsilon > 0$ , find an open subset  $U$ , such that  $A \subset U$  and  $\lambda(U) \leq \epsilon$ .

**Exercise 9.10.** For the sequence  $K_n$  used in the construction of the Cantor set  $K$ , show that  $\lambda(K_{n+1}) = \frac{2}{3}\lambda(K_n)$ . Then use this to show that  $\lambda(K) = 0$ .

**Exercise 9.11.** Suppose  $A \subset [a, b]$  satisfies  $\mu([a, b] - A) = 0$ . Prove that  $A$  is dense in  $[a, b]$ . In other words, for any nonempty open interval  $I$  inside  $[a, b]$ , the intersection  $A \cap I$  is not empty.

**Exercise 9.12.** Prove that if  $A$  and  $B$  are Lebesgue measurable, then  $A \cup B$ ,  $A \cap B$  and  $A - B$  are Lebesgue measurable, and  $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$ .

**Exercise 9.13.** Prove that any compact subset  $K$  is Lebesgue measurable and  $\mu(K) = \lambda(K)$ .

**Exercise 9.14.** Prove that  $\mu_*(\sqcup A_i) \geq \sum \mu_*(A_i)$  for a countable disjoint union. In case  $A_i$  are measurable, further prove that  $\sqcup A_i$  is measurable and  $\mu(\sqcup A_i) = \sum \mu(A_i)$ .

**Exercise 9.15.** Prove that the union and the intersection of countably many uniformly bounded Lebesgue measurable subsets are Lebesgue measurable.

**Exercise 9.16.** Prove that if  $\mu^*(A) = 0$ , then  $\mu^*(A \cup B) = \mu^*(B)$  and  $\mu_*(A \cup B) = \mu_*(B)$ .

**Exercise 9.17.** Suppose  $A$  and  $B$  differ by a subset of measure zero. In other words, we have  $\mu^*(A - B) = \mu^*(B - A) = 0$ . Prove that  $A$  is measurable if and only if  $B$  is measurable. Moreover, we have  $\mu(A) = \mu(B)$  in case both are measurable.

**Exercise 9.18.** Prove that the following are equivalent to the Lebesgue measurability of a subset  $A \subset \mathbb{R}$ .

1. For any  $\epsilon > 0$ , there is open  $U \supset A$ , such that  $\mu^*(U - A) < \epsilon$ .
2. For any  $\epsilon > 0$ , there is closed  $K \subset A$ , such that  $\mu^*(A - K) < \epsilon$ .
3. For any  $\epsilon > 0$ , there is a finite union  $U$  of open intervals, such that  $\mu^*((U - A) \cup (A - U)) < \epsilon$ .
4. For any  $\epsilon > 0$ , there is a finite union  $U$  of open intervals, such that  $\mu^*(U - A) < \epsilon$  and  $\mu^*(A - U) < \epsilon$ .

**Exercise 9.19.** Recall from Exercise 2.42 that a real function  $f$  is *Lipschitz* if there is a constant  $L$ , such that  $|f(x) - f(y)| \leq L|x - y|$  for any  $x$  and  $y$ .

1. Prove that the Lipschitz function satisfies  $\mu^*(f(A)) \leq L\mu^*(A)$ .

2. Prove that if  $A$  is Lebesgue measurable, then  $f(A)$  is also Lebesgue measurable.

Example 11.4.5 shows that a continuous function may not take Lebesgue measurable subsets to Lebesgue measurable subsets.

### 9.3 Outer Measure

The inner measure in  $\mathbb{R}$  is introduced by using the outer measure of the complement. This suggests that we should be able to characterize Lebesgue measurability by using the outer measure only. Once this is achieved, the outer measure and measurability can be extended to general abstract setting.

#### Carathéodory Theorem

The following says that a subset is Lebesgue measurable if and only if it and its complement can be used to “cut” the outer measure of any subset.

**Theorem 9.3.1** (Carathéodory). *A bounded subset  $A \subset \mathbb{R}$  is Lebesgue measurable if and only if*

$$\mu^*(X) = \mu^*(X \cap A) + \mu^*(X - A)$$

*for any bounded subset  $X$ .*

By the third property in Proposition 9.2.2, we always have  $\mu^*(X) \leq \mu^*(X \cap A) + \mu^*(X - A)$ . The theorem says that Lebesgue measurable subsets induce “good cuts”. The proof makes use of the following technical result.

**Lemma 9.3.2.** *If  $A$  and  $B$  are disjoint, then*

$$\mu_*(A \sqcup B) \leq \mu_*(A) + \mu^*(B) \leq \mu^*(A \sqcup B).$$

*Proof.* For any  $\epsilon > 0$ , there are open  $U \supset A \cup B$  and closed  $K \subset A$ , such that

$$\lambda(U) < \mu^*(A \sqcup B) + \epsilon, \quad \lambda(K) > \mu_*(A) - \epsilon.$$

Since  $U - K$  is open and contains  $B$ , we also have  $\lambda(U - K) \geq \mu^*(B)$ . Then

$$\mu^*(A \sqcup B) + \epsilon > \lambda(U) = \lambda(K) + \lambda(U - K) > \mu_*(A) - \epsilon + \mu^*(B).$$

Because  $\epsilon$  is arbitrary, this implies the second inequality of the lemma.

Now we turn to the first inequality. For any  $\epsilon > 0$ , there are closed  $K \subset A \cup B$  and open  $U \supset B$ , such that

$$\lambda(K) > \mu_*(A \cup B) - \epsilon, \quad \lambda(U) < \mu^*(B) + \epsilon.$$

Since  $K - U$  is closed and contained in  $A$ , we also have  $\mu_*(A) \geq \lambda(K - U)$ . By the second part of Exercise 9.7, we have

$$\mu_*(A \cup B) - \epsilon < \lambda(K) \leq \lambda(K - U) + \lambda(U) < \mu_*(A) + \mu^*(B) + \epsilon.$$

Because  $\epsilon$  is arbitrary, this further implies the first inequality of the lemma.  $\square$

*Proof of Theorem 9.3.1.* For the sufficiency, we assume that the “cutting equality” holds. Applying the equality to a bounded interval  $X$  containing  $A$ , we get

$$\begin{aligned}\mu^*(A) + \mu^*(X - A) &= \mu^*(X) && \text{(cutting equality)} \\ &= \mu_*(X) && (X \text{ is Lebesgue measurable}) \\ &\leq \mu_*(A) + \mu^*(X - A). && \text{(Lemma 9.3.2)}\end{aligned}$$

This implies  $\mu^*(A) \leq \mu_*(A)$ , so that  $A$  is Lebesgue measurable.

For the necessity, we first prove the “cutting equality” under the stronger assumption that  $A \cap U$  is Lebesgue measurable for any open  $U$ . For any  $X$  and any  $\epsilon > 0$ , there is open  $U \supset X$ , such that  $\lambda(U) < \mu^*(X) + \epsilon$ . Then

$$\begin{aligned}\mu^*(X) + \epsilon &> \mu^*(U) \geq \mu_*(U \cap A) + \mu^*(U - A) && \text{(Lemma 9.3.2)} \\ &= \mu^*(U \cap A) + \mu^*(U - A) && \text{(stronger assumption)} \\ &\geq \mu^*(X \cap A) + \mu^*(X - A). && \text{(Proposition 9.3.1)}\end{aligned}$$

Since  $\epsilon$  is arbitrary, we get  $\mu^*(X) \geq \mu^*(X \cap A) + \mu^*(X - A)$ . Since  $\mu^*(X) \leq \mu^*(X \cap A) + \mu^*(X - A)$  always holds by the third property in Proposition 9.2.2, we get the cutting equality.

For the general case of the necessity, we note that the intersection of any two open subsets is open. By Example 9.2.3, therefore, open subsets satisfy the stronger assumption. For Lebesgue measurable  $A$  and open  $U$ , we then have

$$\begin{aligned}\mu^*(A \cap U) + \mu^*(A - U) &= \mu^*(A) && (U \text{ gives cutting equality}) \\ &= \mu_*(A) && (A \text{ is Lebesgue measurable}) \\ &\leq \mu_*(A \cap U) + \mu^*(A - U). && \text{(Lemma 9.3.2)}\end{aligned}$$

Therefore  $\mu^*(A \cap U) \leq \mu_*(A \cap U)$ . This proves that  $A \cap U$  is Lebesgue measurable for any open  $U$ . Then by the special case proved above,  $A$  gives the cutting equality.  $\square$

## Abstract Outer Measure

Theorem 9.3.1 suggests how to establish a general measure theory by using outer measure only. First, Proposition 9.2.2 suggests the definition of general outer measure.

**Definition 9.3.3.** An *outer measure* on a set  $X$  is an extended number  $\mu^*(A)$  assigned to each subset  $A \subset X$ , such that the following are satisfied.

1. *Empty Set:*  $\mu^*(\emptyset) = 0$ .
2. *Monotone:*  $A \subset B$  implies  $\mu^*(A) \leq \mu^*(B)$ .
3. *Countable Subadditivity:*  $\mu^*(\cup A_i) \leq \sum \mu^*(A_i)$  for countable union  $\cup A_i$ .

By extended number, we mean that the value can be  $+\infty$ . The first two properties imply that  $\mu^*(A) \geq 0$  for any  $A$ .

**Example 9.3.1.** A trivial example of the outer measure is  $\mu^*(A) = 1$  for any nonempty subset  $A$ , and  $\mu^*(\emptyset) = 0$ . More generally, we may fix  $Y \subset X$  and define

$$\mu^*(A) = \begin{cases} 1, & \text{if } A \not\subset Y, \\ 0, & \text{if } A \subset Y. \end{cases}$$

**Example 9.3.2.** Let  $X$  be any set. For any  $A \subset X$ , let  $\mu^*(A)$  be the number of elements in  $A$ . In case  $A$  is infinite, let  $\mu^*(A) = +\infty$ . Then  $\mu^*$  is an outer measure.

**Example 9.3.3.** The *Lebesgue outer measure* on  $\mathbb{R}$  has been defined for bounded subsets. The definition can be extended to any subset of  $\mathbb{R}$  by taking the limit of the bounded case

$$\mu^*(A) = \lim_{r \rightarrow +\infty} \mu^*(A \cap [-r, r]) = \sup_{r > 0} \mu^*(A \cap [-r, r]).$$

Alternatively, we can keep using the old definition

$$\mu^*(A) = \inf\{\lambda(U) : A \subset U, U \text{ open}\}.$$

Here we still express (not necessarily bounded)  $U$  as a disjoint union of open intervals and define  $\lambda(U)$  as the sum of the lengths of these open intervals.

**Exercise 9.20.** Prove that the definition of the outer measure is not changed if we additionally require that union in the third condition to be disjoint.

**Exercise 9.21.** Prove that the sum of countably many outer measures is an outer measure. Prove that the positive scalar multiple of an outer measure is an outer measure.

**Exercise 9.22.** Suppose  $\mu^*$  is an outer measure on  $X$  and  $Y \subset X$ . Prove that the restriction of  $\mu^*$  to subsets of  $Y$  is an outer measure on  $Y$ .

**Exercise 9.23.** Suppose  $\mu_1^*$  and  $\mu_2^*$  are outer measures on  $X_1$  and  $X_2$ . Prove that  $\mu^*(A) = \mu_1^*(A \cap X_1) + \mu_2^*(A \cap X_2)$  is an outer measure on  $X_1 \sqcup X_2$ . Extend the result to any disjoint union.

**Exercise 9.24.** Prove that the first definition in Example 9.3.3 indeed gives an outer measure.

**Exercise 9.25.** Prove that the extended definition of  $\lambda(U)$  in Example 9.3.3 satisfies  $\lambda(\cup U_i) \leq \sum \lambda(U_i)$  for countable unions  $\cup U_i$  of open subsets. Then use this to show that the second definition in Example 9.3.3 gives an outer measure.

In general, suppose  $\mathcal{C}$  is a collection of subsets in  $X$  and  $\lambda$  is a non-negative valued function on  $\mathcal{C}$ . What is the condition on  $\lambda$  so that  $\mu^*(A) = \inf\{\lambda(C) : A \subset C, C \in \mathcal{C}\}$  defines an outer measure?

**Exercise 9.26.** Prove that the two definitions in Example 9.3.3 give equal outer measures.

## Measurability with Respect to Outer Measure

Theorem 9.3.1 suggests the concept of measurability in general setting.

**Definition 9.3.4.** Given an outer measure  $\mu^*$  on a set  $X$ , a subset  $A$  is *measurable* (with respect to  $\mu^*$ ) if

$$\mu^*(Y) = \mu^*(Y \cap A) + \mu^*(Y - A), \text{ for any } Y \subset X. \quad (9.3.1)$$

The *measure* of the measurable subset is  $\mu(A) = \mu^*(A)$ .

By the subadditivity of outer measure, we always have

$$\mu^*(Y) \leq \mu^*(Y \cap A) + \mu^*(Y - A).$$

Therefore the equality (9.3.1) always holds when  $\mu^*(Y) = +\infty$ , and the equality (9.3.1) is really equivalent to

$$\mu^*(Y) \geq \mu^*(Y \cap A) + \mu^*(Y - A), \text{ for any } Y \subset X, \mu^*(Y) < +\infty. \quad (9.3.2)$$

Denote  $B = X - A$ . Then  $X = A \sqcup B$  is a partition of  $X$  into disjoint subsets, and the condition (9.3.1) becomes

$$\mu^*(Y) = \mu^*(Y \cap A) + \mu^*(Y \cap B), \text{ for any } Y \subset X.$$

This shows that the condition is symmetric with respect to  $A$  and  $B$ . In particular, this shows that  $A$  is measurable if and only if  $X - A$  is measurable. Moreover, it is easy to show by induction that if  $X = A_1 \sqcup \cdots \sqcup A_n$  for disjoint measurable subsets  $A_1, \dots, A_n$ , then for any subset  $Y$ , we have

$$\mu^*(Y) = \mu^*(Y \cap A_1) + \cdots + \mu^*(Y \cap A_n). \quad (9.3.3)$$

**Proposition 9.3.5.** *The countable unions and intersections of measurable subsets are measurable. The subtraction of two measurable subsets is measurable. Moreover, the measure has the following properties.*

1. If  $\mu^*(A) = 0$ , then any subset  $B \subset A$  is measurable and  $\mu(B) = 0$ .
2.  $\mu(A) \geq 0$ .
3. If countably many  $A_i$  are measurable and disjoint, then  $\mu(\sqcup A_i) = \sum \mu(A_i)$ .

The first property actually means two properties. First,  $\mu^*(A) = 0$  implies  $A$  is measurable and  $\mu(A) = 0$ . Second,  $\mu(A) = 0$  and  $B \subset A$  imply  $B$  is measurable and  $\mu(B) = 0$ .

*Proof.* Suppose  $\mu^*(A) = 0$  and  $B \subset A$ . Then for any  $Y$ , we have

$$0 \leq \mu^*(Y \cap B) \leq \mu^*(B) \leq \mu^*(A) = 0.$$

Therefore  $\mu^*(Y \cap B) = 0$  and

$$\mu^*(Y) \leq \mu^*(Y \cap B) + \mu^*(Y - B) = \mu^*(Y - B) \leq \mu^*(Y).$$

This verifies the condition (9.3.1) for  $B$ .

The positivity  $\mu(A) \geq 0$  follows from  $\mu^*(A) \geq 0$  for any  $A$ .

For measurable  $A, B$  and any  $Y$ , we have

$$\begin{aligned}\mu^*(Y) &= \mu^*(Y \cap A) + \mu^*(Y - A) \\ &= \mu^*(Y \cap A) + \mu^*((Y - A) \cap B) + \mu^*(Y - A - B) \\ &= \mu^*(Y \cap (A \cup B)) + \mu^*(Y - A \cup B).\end{aligned}$$

In the first equality, we use  $A$  to cut  $Y$ . In the second equality, we use  $B$  to cut  $Y - A$ . In the third equality, we use  $A$  to cut  $Y \cap (A \cup B)$ . This proves that  $A \cup B$  is measurable. Similar reason shows that  $A \cap B$  is measurable. Moreover, the measurability of  $B$  implies the measurability of  $X - B$ , so that the subtraction  $A - B = A \cap (X - B)$  is measurable.

When  $A$  and  $B$  are disjoint and measurable, we use  $A$  to cut  $A \sqcup B$  and get

$$\mu(A \sqcup B) = \mu^*((A \sqcup B) \cap A) + \mu^*(A \sqcup B - A) = \mu(A) + \mu(B).$$

This finishes the proof for the finite case of the proposition.

The countable case will be proved as the limit of finite case. First consider measurable and disjoint  $A_i, i \in \mathbb{N}$ . The finite unions  $B_n = \sqcup_{i=1}^n A_i$  are also measurable, and we have

$$\mu^*(Y) = \mu^*(Y \cap B_n) + \mu^*(Y - B_n)$$

for any subset  $Y$ . The union  $A = \sqcup A_i = \cup B_n$  is the “limit” of “increasing sequence”  $B_n$ . We wish to prove that  $A$  is measurable, which means the inequality (9.3.2). We expect to prove the inequality as the limit of the equality for  $B_n$ .

By  $B_n \subset A$ , we have  $\mu^*(Y - A) \leq \mu^*(Y - B_n)$ . So it remains to show that  $\mu^*(Y \cap A)$  is not much bigger than  $\mu^*(Y \cap B_n)$ . The difference can be computed by using the measurable  $B_n$  to cut  $Y \cap A$

$$\begin{aligned}\mu^*(Y \cap A) &= \mu^*(Y \cap A \cap B_n) + \mu^*(Y \cap A - B_n) \\ &= \mu^*(Y \cap B_n) + \mu^*(Y \cap (A - B_n)) \\ &\leq \mu^*(Y \cap B_n) + \sum_{i>n} \mu^*(Y \cap A_i),\end{aligned}$$

where the inequality is due to  $A - B_n = \sqcup_{i>n} A_i$  and the countable subadditivity of  $\mu^*$ . Then we get

$$\begin{aligned}\mu^*(Y) &= \mu^*(Y \cap B_n) + \mu^*(Y - B_n) \\ &\geq \mu^*(Y \cap A) - \sum_{i>n} \mu^*(Y \cap A_i) + \mu^*(Y - A).\end{aligned}$$

To get the inequality (9.3.2), it remains to prove  $\lim_{n \rightarrow \infty} \sum_{i>n} \mu^*(Y \cap A_i) = 0$ .

In case  $\mu^*(Y) < +\infty$ , we may use  $X = A_1 \sqcup \cdots \sqcup A_n \sqcup (X - B_n)$  in (9.3.3) to get

$$\mu^*(Y \cap A_1) + \cdots + \mu^*(Y \cap A_n) = \mu^*(Y \cap B_n) \leq \mu^*(Y) < +\infty. \quad (9.3.4)$$

So for fixed  $Y$ , the partial sums  $\sum_{i>n} \mu^*(Y \cap A_i)$  are bounded. This implies the convergence of the series  $\sum \mu^*(Y \cap A_i)$ , so that  $\lim_{n \rightarrow \infty} \sum_{i>n} \mu^*(Y \cap A_i) = 0$ . This proves that  $A$  is measurable.

Taking  $Y = A$  in (9.3.4), we get  $\sum \mu(A_i) \leq \mu(A)$ . On the other hand, we also have  $\sum \mu(A_i) \geq \mu(A)$  by the countable subadditivity. Therefore the equality in the third property is proved for the case  $\mu(A)$  is finite. The case of infinite  $\mu(A)$  follows directly from the countable subadditivity.

Finally, suppose  $A_i$  are measurable (but not necessarily disjoint). Then  $C_i = A_i - A_1 \cup \dots \cup A_{i-1}$  are measurable and disjoint, so that the union  $\cup A_i = \sqcup C_i$  is measurable. The measurability of the intersection  $\cap A_i$  follows from  $\cap A_i = X - \cup(X - A_i)$ .  $\square$

**Example 9.3.4.** Consider the outer measure in Example 9.3.1. The only measurable subsets are  $\emptyset$  and  $X$ . In the more general case, a subset  $A$  is measurable if and only if  $A - Y$  is either  $\emptyset$  or  $X - Y$ . This means that either  $A \supset X - Y$  or  $A \subset Y$ .

**Example 9.3.5.** With respect to the outer measure in Example 9.3.2. Any subset is measurable. The measure is again the number of elements.

**Example 9.3.6.** Consider the Lebesgue outer measure in Example 9.3.3. If a bounded subset  $A$  is measurable according to Definition 9.3.4, then by restricting 9.3.1 to bounded subsets and using Theorem 9.3.1, we see that  $A$  is measurable according to Definition 9.2.1.

Conversely, if  $A$  is measurable in the sense of Definition 9.2.1, then (9.3.1) is satisfied for bounded  $Y$ . Therefore for any  $Y$ , (9.3.1) is satisfied for  $Y \cap [-r, r]$ . Taking the limit of (9.3.1) as  $r \rightarrow +\infty$ , we see that (9.3.1) is satisfied for  $Y$ .

We conclude that for a bounded subset of  $\mathbb{R}$ , Definitions 9.2.1 and 9.3.4 are equivalent. For an unbounded subset  $A$ , we may use  $A = \cup_{n=1}^{+\infty} (A \cap [-n, n])$  and Proposition 9.3.5 to conclude that  $A$  is Lebesgue measurable if and only if for any  $r > 0$ ,  $A \cap [-r, r]$  is measurable in the sense of Definition 9.2.1.

**Exercise 9.27.** For a disjoint union  $X = A_1 \sqcup \dots \sqcup A_n$  of measurable subsets, prove (9.3.3).

**Exercise 9.28.** Suppose  $\mu_1^*$  and  $\mu_2^*$  are outer measures.

1. Prove that  $\mu_1^* + \mu_2^*$  is an outer measure.
2. Prove that if a subset is  $\mu_1^*$ -measurable and  $\mu_2^*$ -measurable, then the subset is  $(\mu_1^* + \mu_2^*)$ -measurable.
3. Suppose  $\mu_1^*(X), \mu_2^*(X) < +\infty$ . Prove that if a subset is  $(\mu_1^* + \mu_2^*)$ -measurable, then the subset is  $\mu_1^*$ -measurable and  $\mu_2^*$ -measurable.
4. Show that the finite assumption in the third part is necessary.

**Exercise 9.29.** Suppose  $\mu^*$  is an outer measure on  $X$  and  $Y \subset X$ . Exercise 9.22 says that the restriction of  $\mu^*$  to subsets of  $Y$  is an outer measure on  $Y$ .

1. Prove that if  $Y$  is measurable with respect to  $\mu^*$ , then a subset of  $Y$  is measurable with respect to  $\mu_Y^*$  if and only if it is measurable with respect to  $\mu^*$ .
2. What may happen in the second part if  $Y$  is not measurable with respect to  $\mu^*$ ?

**Exercise 9.30.** Suppose  $\mu_1^*$  and  $\mu_2^*$  are outer measures on  $X_1$  and  $X_2$ . Exercise 9.23 says that  $\mu^*(A) = \mu_1^*(A \cap X_1) + \mu_2^*(A \cap X_2)$  is an outer measure on  $X_1 \sqcup X_2$ . Prove that  $A$  is measurable with respect to  $\mu^*$  if and only if  $A \cap X_1$  is measurable with respect to  $\mu_1^*$  and  $A \cap X_2$  is measurable with respect to  $\mu_2^*$ .

**Exercise 9.31.** Suppose  $\mu^*$  is an outer measure on  $X$  invariant under an invertible map  $\phi: X \rightarrow X$ ,

$$\mu^*(\phi(A)) = \mu^*(A).$$

Prove that  $A$  is measurable with respect to  $\mu^*$  if and only if  $\phi(A)$  is measurable, and  $\mu(\phi(A)) = \mu(A)$ . Then prove that the Lebesgue measure satisfies

$$\mu(aA + b) = |a|\mu(A), \quad aA + b = \{ax + b: x \in A\}.$$

## 9.4 Measure Space

Proposition 9.3.5 summarizes the measure theory induced from an outer measure. We may take the second and third properties in the proposition as the general definition of measure theory. However, the properties only apply to measurable subsets. Therefore before talking about the properties, we need to describe the collection of all measurable subsets. This is the concept of  $\sigma$ -algebra.

### $\sigma$ -Algebra

Inspired by the set theoretical part of Proposition 9.3.5, we introduce the following concept.

**Definition 9.4.1.** A  $\sigma$ -algebra on a set  $X$  is a collection  $\Sigma$  of subsets of  $X$ , such that the following are satisfied.

1.  $X \in \Sigma$ .
2.  $A, B \in \Sigma \implies A - B \in \Sigma$ .
3. Countably many  $A_i \in \Sigma \implies \cup A_i \in \Sigma$ .

The subsets in  $\Sigma$  are called *measurable*.

The subtraction and countable union of measurable subsets are measurable. By  $\cap A_i = A_1 - \cup(A_1 - A_i)$ , the countable intersection of measurable subsets is also measurable.

The Lebesgue  $\sigma$ -algebra on  $\mathbb{R}$  is the collection of all Lebesgue measurable subsets, including the unbounded measurable subsets as extended in Examples 9.3.3 and 9.3.6.

**Example 9.4.1.** On any set  $X$ , the smallest  $\sigma$ -algebra is  $\{\emptyset, X\}$ , and the biggest  $\sigma$ -algebra is the collection  $\mathcal{P}(X)$  of all subsets in  $X$ .



**Example 9.4.2.** It is easy to see that the intersection of  $\sigma$ -algebras is still a  $\sigma$ -algebra. Therefore we may start with any collection  $\mathcal{C}$  of subsets of  $X$  and take the intersection of all  $\sigma$ -algebras on  $X$  that contain  $\mathcal{C}$ . The result is the smallest  $\sigma$ -algebra such that any subset in  $\mathcal{C}$  is measurable.

The smallest  $\sigma$ -algebra on  $X$  such that every single point is measurable consists of those  $A \subset X$  such that either  $A$  or  $X - A$  is countable.

**Example 9.4.3 (Borel Set).** The *Borel  $\sigma$ -algebra* is the smallest  $\sigma$ -algebra on  $\mathbb{R}$  such that every open interval is measurable. The subsets in the Borel  $\sigma$ -algebra are *Borel sets*.

Let  $\epsilon_n > 0$  converge to 0. Then we have

$$[a, b] = \bigcup (a - \epsilon_n, b + \epsilon_n), \quad (a, b] = \bigcap (a, b + \epsilon_n), \quad [a, b) = \bigcap (a - \epsilon_n, b).$$

Therefore all intervals are Borel sets. By Theorem 6.4.6, all open subsets are Borel sets. Taking the complement, all closed subsets are also Borel sets.

Since open intervals are Lebesgue measurable, the Lebesgue  $\sigma$ -algebra contains all open intervals. Therefore the Lebesgue  $\sigma$ -algebra contains the Borel  $\sigma$ -algebra. In other words, all Borel sets are Lebesgue measurable.

**Exercise 9.32.** Suppose there are countably many measurable subsets  $X_i \subset X$ , such that  $X = \bigcup X_i$ . Prove that  $A \subset X$  is measurable if and only if each  $A \cap X_i$  is measurable.

**Exercise 9.33.** Prove that the third condition in the definition of  $\sigma$ -algebra can be replaced by the following: If countably many  $A_i \in \Sigma$  are disjoint, then  $\bigcup A_i \in \Sigma$ .

**Exercise 9.34.** Prove that a collection  $\Sigma$  of subsets of  $X$  is a  $\sigma$ -algebra if and only if the following are satisfied.

1.  $X \in \Sigma$ .
2.  $A, B \in \Sigma \implies A \cup B, A - B \in \Sigma$ .
3.  $A_i \in \Sigma, A_i \subset A_{i+1} \implies \bigcup A_i \in \Sigma$ .

**Exercise 9.35.** Prove that the smallest  $\sigma$ -algebra containing each of the following collections is the Borel  $\sigma$ -algebra.

1. All closed intervals.
2. All left open and right closed intervals.
3. All intervals.
4. All intervals of the form  $(a, +\infty)$ .
5. All intervals of length  $< 1$ .
6. All open subsets.
7. All closed subsets.
8. All compact subsets.

**Exercise 9.36.** Suppose  $\Sigma$  is a  $\sigma$ -algebra on  $X$  and  $f: X \rightarrow Y$  is a map. Prove that the *push forward*

$$f_*(\Sigma) = \{B \in Y: f^{-1}(B) \in \Sigma\}$$

is a  $\sigma$ -algebra on  $Y$ . However, the *image*

$$f(\Sigma) = \{f(A) : A \in \Sigma\}$$

may not be a  $\sigma$ -algebra, even when  $f$  is onto.

**Exercise 9.37.** Suppose  $\Sigma$  is a  $\sigma$ -algebra on  $Y$  and  $f: X \rightarrow Y$  is a map. Prove that the *preimage*

$$f^{-1}(\Sigma) = \{f^{-1}(B) : B \in \Sigma\}$$

is a  $\sigma$ -algebra on  $X$ . However, the *pull back*

$$f^*(\Sigma) = \{A \subset X : f(A) \in \Sigma\}$$

may not be a  $\sigma$ -algebra.

## Abstract Measure

The numerical aspect of Proposition 9.3.5 leads to the following concept.

**Definition 9.4.2.** A *measure* on a  $\sigma$ -algebra  $\Sigma$  assigns an extended number  $\mu(A)$  to each  $A \in \Sigma$ , such that the following are satisfied.

1. *Empty:*  $\mu(\emptyset) = 0$ .
2. *Positivity:*  $\mu(A) \geq 0$ .
3. *Countable Additivity:* If countably many  $A_i \in \Sigma$  are disjoint, then  $\mu(\sqcup A_i) = \sum \mu(A_i)$ .

A *measure space*  $(X, \Sigma, \mu)$  consists of a set  $X$ , a  $\sigma$ -algebra  $\Sigma$  on  $X$  and a measure  $\mu$  on  $\Sigma$ .

The following is a useful condition because many results that hold under the condition  $\mu(X) < +\infty$  can be extended under the following condition.

**Definition 9.4.3.** A subset is  *$\sigma$ -finite* if it is contained in a union of countably many subsets of finite measure. A measure is  *$\sigma$ -finite* if the whole space is  $\sigma$ -finite.

A measure space can be induced by an outer measure. Of course a measure space can also be constructed by any other method, as long as the properties are satisfied.

**Example 9.4.4.** For any set  $X$ , let  $\Sigma$  be the power set  $\mathcal{P}(X)$  of all subsets in  $X$ . For finite  $A \subset X$ , define  $\mu(A)$  to be the number of elements in  $A$ . For infinite  $A \subset X$ , define  $\mu(A) = +\infty$ . Then we get the *counting measure*. This is the measure in Example 9.3.5 and is induced by the outer measure in Example 9.3.2.

The counting measure is  $\sigma$ -finite if and only if  $X$  is countable.

**Example 9.4.5.** For any set  $X$ , we still take  $\Sigma$  to be the power set  $\mathcal{P}(X)$ . We fix an element  $a \in X$  and define  $\mu(A) = 1$  when  $a \in A$  and  $\mu(A) = 0$  when  $a \notin A$ . Then we get the *Dirac measure* concentrated at  $a$ .

**Example 9.4.6.** The *Lebesgue measure space* on  $\mathbb{R}$  is the one in Example 9.3.6. Specifically, a bounded subset  $A$  is measurable if and only if the inner and outer measures are equal (Definition 9.2.1). An unbounded subset  $A$  is measurable if and only if the bounded subset  $A \cap [-r, r]$  is measurable for any  $r > 0$ . This is also the same as  $A \cap [-n, n]$  being measurable for any  $n \in \mathbb{N}$ . The Lebesgue measure of any measurable subset may be computed from the bounded ones by

$$\mu(A) = \lim_{r \rightarrow +\infty} \mu(A \cap [-r, r]) = \sup_{r > 0} \mu(A \cap [-r, r]).$$

The Lebesgue measure is  $\sigma$ -finite.

**Exercise 9.38.** Suppose  $X$  is any set, and a non-negative number  $\mu_x$  is assigned to any  $x \in X$ . For any  $A \subset X$ , define

$$\mu(A) = \begin{cases} \sum_{x \in A} \mu_x, & \text{if } \mu_x \neq 0 \text{ for only countably many } x \in A, \\ +\infty, & \text{if } \mu_x \neq 0 \text{ for uncountably many } x \in A. \end{cases}$$

Verify that  $\mu$  is a measure on the collection of all subsets in  $X$ . When is the measure  $\sigma$ -finite?

**Exercise 9.39.** Prove that the sum of countably many measures on the same  $\sigma$ -algebra is a measure. If each measure is  $\sigma$ -finite, is the sum  $\sigma$ -finite?

**Exercise 9.40.** Prove that a union of countably many  $\sigma$ -finite subsets is still  $\sigma$ -finite.

**Exercise 9.41.** Suppose  $(X, \Sigma, \mu)$  is a measure space and  $Y \in \Sigma$ . Prove that  $\Sigma_Y = \{A \subset Y : A \in \Sigma\}$  is a  $\sigma$ -algebra on  $Y$ , and the restriction  $\mu_Y$  of  $\mu$  to  $\Sigma_Y$  is a measure. What may happen if  $Y \notin \Sigma$ ?

**Exercise 9.42.** Suppose  $(X_1, \Sigma_1, \mu_1)$  and  $(X_2, \Sigma_2, \mu_2)$  are measure spaces.

1. Prove that  $\Sigma_1 \sqcup \Sigma_2 = \{A_1 \sqcup A_2 : A_1 \in \Sigma_1, A_2 \in \Sigma_2\}$  is a  $\Sigma$ -algebra on  $X_1 \sqcup X_2$ .
2. Prove that  $\mu(A_1 \sqcup A_2) = \mu_1(A_1) + \mu_2(A_2)$  is a measure on  $X_1 \sqcup X_2$ .
3. Suppose  $(X, \Sigma, \mu)$  is a measure space, and  $Y \in \Sigma$ . Prove that  $(X, \Sigma, \mu)$  is the disjoint union of the restricted measure spaces on  $Y$  and  $X - Y$  constructed in Exercise 9.41.

Extend the result to any disjoint union.

**Proposition 9.4.4.** *A measure has the following properties.*

1. *Monotone:*  $A \subset B$  implies  $\mu(A) \leq \mu(B)$ .
2. *Countable Subadditivity:*  $\mu(\cup A_i) \leq \sum \mu(A_i)$  for a countable union  $\cup A_i$ .
3. *Monotone Limit:* If  $A_i \subset A_{i+1}$  for all  $i$ , then  $\mu(\cup A_i) = \lim \mu(A_i)$ . If  $A_i \supset A_{i+1}$  for all  $i$  and  $\mu(A_1)$  is finite, then  $\mu(\cap A_i) = \lim \mu(A_i)$ .

*Proof.* For measurable  $A \subset B$ , we have

$$\mu(B) = \mu(A) + \mu(B - A) \geq \mu(A),$$

where the equality is by the additivity and the inequality is by the positivity.

For measurable  $A_i$ , the subsets  $C_i = A_i - A_1 \cup \cdots \cup A_{i-1}$  are measurable and disjoint. Then  $\cup A_i = \sqcup C_i$  and

$$\mu(\cup A_i) = \sum \mu(C_i) \leq \sum \mu(A_i),$$

where the equality is by the countable additivity and the inequality is by the monotone property that was just proved.

If  $A_i \subset A_{i+1}$ , then denote  $A_0 = \emptyset$  and by the countable additivity,

$$\begin{aligned} \mu(\cup A_i) &= \mu(\sqcup_{i=1}^{\infty} (A_i - A_{i-1})) = \sum_{i=1}^{\infty} \mu(A_i - A_{i-1}) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(A_i - A_{i-1}) = \lim_{n \rightarrow \infty} \mu(\sqcup_{i=1}^n (A_i - A_{i-1})) = \lim_{n \rightarrow \infty} \mu(A_n). \end{aligned}$$

If  $A_i \supset A_{i+1}$ , then  $B_i = A_1 - A_i \subset B_{i+1} = A_1 - A_{i+1}$ . Since  $\mu(A_1)$  is finite, we have

$$\mu(\cup B_i) = \mu(A_1 - \cap A_i) = \mu(A_1) - \mu(\cap A_i), \quad \mu(B_i) = \mu(A_1) - \mu(A_i).$$

By  $\mu(\cup B_i) = \lim_{n \rightarrow \infty} \mu(B_i)$  for the increasing sequence  $B_i$ , we get  $\mu(\cap A_i) = \lim_{n \rightarrow \infty} \mu(A_i)$  for the decreasing sequence  $A_i$ .  $\square$

**Exercise 9.43.** Show that the condition  $\mu(A_1) < +\infty$  cannot be removed from the third property in Proposition 9.4.4.

**Exercise 9.44.** Prove that a measurable subset  $A$  is  $\sigma$ -finite if and only if it is the union of countably many disjoint measurable subsets of finite measure.

**Exercise 9.45.** Prove that a measurable subset  $A$  is  $\sigma$ -finite if and only if it is the union of an increasing sequence of measurable subsets of finite measure.

## Subset of Measure Zero

Subsets of measure zero can be considered as negligible for many purposes. This means that the difference within a subset of measure zero often do not affect the outcome.

**Definition 9.4.5.** Let  $(X, \Sigma, \mu)$  be a measure space. If there is a subset  $A \in \Sigma$  of measure zero, such that a property related to points  $x \in X$  holds for any  $x \notin A$ , then we say the property holds *almost everywhere*.

For example, if  $f(x) = g(x)$  for all  $x$  out of a subset of measure zero, then  $f(x) = g(x)$  almost everywhere. Specifically, the Dirichlet function is zero almost everywhere. For another example, two subsets  $A$  and  $B$  are *almost the same* if the difference

$$A \Delta B = (A - B) \cup (B - A)$$

is contained in a subset in  $\Sigma$  of measure zero.

**Proposition 9.4.6.** *The subsets of measure zero have the following properties.*

1. If  $A, B \in \Sigma$ ,  $B \subset A$ , and  $\mu(A) = 0$ , then  $\mu(B) = 0$ .
2. If  $A, B \in \Sigma$  and the difference  $(A - B) \cup (B - A)$  has measure zero, then  $\mu(A) = \mu(B)$ .
3. A countable union of subsets of measure zero has measure zero.
4. If countably many  $A_i \in \Sigma$  satisfy  $\mu(A_i \cap A_j) = 0$  for any  $i \neq j$ , then  $\mu(\cup A_i) = \sum \mu(A_i)$ .

*Proof.* For the first statement, we have

$$0 \leq \mu(A) \leq \mu(B) = 0,$$

where the first inequality is by the positivity and the second by the monotone property.

For the second statement, by  $\mu((A - B) \cup (B - A)) = 0$ ,  $A - B \subset (A - B) \cup (B - A)$ , and the first statement, we have  $\mu(A - B) = 0$ . Therefore by  $A \subset (A - B) \cup B$ , we have

$$\mu(A) \leq \mu(A - B) + \mu(B) = \mu(B).$$

By the same reason, we get  $\mu(B) \leq \mu(A)$ .

In the third statement, for countably many subsets  $A_i$  of measure zero, by the positivity and the countable subadditivity, we have

$$0 \leq \mu(\cup A_i) \leq \sum \mu(A_i) = \sum 0 = 0.$$

In the fourth statement, let

$$\begin{aligned} B_i &= A_i - A_1 - A_2 - \cdots - A_{i-1} \\ &= A_i - (A_1 \cap A_i) \cup (A_2 \cap A_i) \cup \cdots \cup (A_{i-1} \cap A_i). \end{aligned}$$

Then  $\cup A_i = \sqcup B_i$ , and by the second part,  $\mu(A_i) = \mu(B_i)$ . Therefore

$$\mu(\cup A_i) = \mu(\sqcup B_i) = \sum \mu(B_i) = \sum \mu(A_i). \quad \square$$

**Exercise 9.46.** Suppose  $f = g$  almost everywhere. Prove that  $f^{-1}(A)$  and  $g^{-1}(A)$  are almost the same for any  $A \subset \mathbb{R}$ .

**Exercise 9.47.** Prove properties of functions that are equal almost everywhere.

1. If  $f = g$  almost everywhere, and  $g = h$  almost everywhere, then  $f = h$  almost everywhere.
2. If  $f = g$  almost everywhere, then  $h \circ f = h \circ g$  almost everywhere. What about  $f \circ h$  and  $g \circ h$ ?
3. If  $f_1 = g_1$  and  $f_2 = g_2$  almost everywhere, then  $f_1 + f_2 = g_1 + g_2$  and  $f_1 f_2 = g_1 g_2$  almost everywhere.
4. If  $f_i = g_i$  almost everywhere for each  $i \in \mathbb{N}$ , then  $\sup f_i = \sup g_i$ ,  $\inf f_i = \inf g_i$ ,  $\overline{\lim}_{i \rightarrow \infty} f_i = \overline{\lim}_{i \rightarrow \infty} g_i$ , and  $\underline{\lim}_{i \rightarrow \infty} f_i = \underline{\lim}_{i \rightarrow \infty} g_i$  almost everywhere.

**Exercise 9.48.** Prove properties of almost the same subsets.

1. If  $A$  and  $B$  are almost the same, and  $B$  and  $C$  are almost the same, then  $A$  and  $C$  are almost the same.
2. If  $A$  and  $B$  are almost the same, and  $C$  and  $D$  are almost the same, then  $A - C$  and  $B - D$  are almost the same.
3. If countably many pairs of  $A_i$  and  $B_i$  are almost the same, then  $\cup_i A_i$  and  $\cup_i B_i$  are almost the same, and  $\cap_i A_i$  and  $\cap_i B_i$  are almost the same.

**Exercise 9.49.** A sequence  $A_i$  is almost increasing if for each  $i$ ,  $A_i - A_{i+1}$  is contained in a measurable subset of measure zero. Prove that if  $A_i$  are measurable and almost increasing, then  $\mu(\cup A_i) = \lim \mu(A_i)$ . What about almost decreasing sequence of measurable subsets?

**Exercise 9.50 (Borel-Cantelli Lemma).** Suppose  $\sum \mu(A_n) < +\infty$ . Prove that almost all  $x \in X$  lie in finitely many  $A_n$ . Moreover, show that the conclusion fails if the condition is relaxed to  $\lim \mu(A_n) = 0$ .

## Complete Measure

The definition of measure takes only the second and third properties from Proposition 9.3.5. If we also include the first property, then we get the following.

**Definition 9.4.7.** A measure space  $(X, \Sigma, \mu)$  is *complete* if

$$\mu(A) = 0, B \subset A \implies B \in \Sigma.$$

By Proposition 9.4.6, the subset  $B$  must also have measure 0. If a measure space is complete, then we only need to talk about subset of measure zero, instead of subset of a (measurable) subset of measure zero.

**Example 9.4.7.** The first property in Proposition 9.3.5 shows that any measure induced by an outer measure is complete. In particular, the Lebesgue measure on  $\mathbb{R}$  is complete, because it is induced from the outer measure. Moreover, the counting measure in Example 9.4.4 and the Dirac measure in Example 9.4.5 are also complete.

**Example 9.4.8.** A simple non-complete measure is given by

$$X = \{1, 2, 3\}, \quad \Sigma = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}, \quad \mu(A) = \begin{cases} 1, & \text{if } 1 \in A, \\ 0, & \text{if } 1 \notin A. \end{cases}$$

We have  $\mu\{2, 3\} = 0$ , but the subsets  $\{2\}$  and  $\{3\}$  are not in  $\Sigma$ .

**Exercise 9.51.** Suppose two subsets  $A$  and  $B$  in a complete measure space are almost the same. Prove that  $A$  is measurable if and only if  $B$  is measurable. Another interpretation of the result is that, in a complete measure space, any almost measurable subset is measurable.

**Exercise 9.52.** Prove that in a complete measure space, a subset  $A$  is measurable if and only if for any  $\epsilon > 0$ , there are measurable subsets  $B$  and  $C$ , such that  $B \subset A \subset C$  and  $\mu(C - B) < \epsilon$ .

**Exercise 9.53.** Prove that in a complete measure space, a subset  $A$  is measurable if and only if for any  $\epsilon > 0$ , there is a measurable subset  $B$ , such that  $(A - B) \cup (B - A)$  is contained in a measurable subset of measure  $\epsilon$ .

To make a measure space  $(X, \Sigma, \mu)$  complete, we must enlarge the  $\sigma$ -algebra to include any subset that is almost the same as a measurable subset. Therefore we introduce

$$\bar{\Sigma} = \{A : A \Delta B \subset C \in \Sigma \text{ for some } B, C \in \Sigma \text{ satisfying } \mu(C) = 0\},$$

and  $\bar{\mu}(A) = \mu(B)$ .

**Proposition 9.4.8.** Suppose  $(X, \Sigma, \mu)$  is a measure space. Then  $\bar{\Sigma}$  is a  $\sigma$ -algebra,  $\bar{\mu}$  is a well defined measure on  $\bar{\Sigma}$ , and  $(X, \bar{\Sigma}, \bar{\mu})$  is a complete measure space.

The measure space  $(X, \bar{\Sigma}, \bar{\mu})$  is called the *completion* of  $(X, \Sigma, \mu)$ . It is the smallest complete measure space containing  $(X, \Sigma, \mu)$ .

*Proof.* We could imagine the difference  $A \Delta B$  as the distance  $d(A, B)$  between subsets  $A$  and  $B$ . We expect  $d(A_1 - A_2, B_1 - B_2) \leq d(A_1, B_1) + d(A_2, B_2)$ . Translated into set theory, this is  $(A_1 - A_2) \Delta (B_1 - B_2) \subset (A_1 \Delta B_1) \cup (A_2 \Delta B_2)$ , which can be verified directly. The set theoretical inclusion implies that if  $A_1, A_2 \in \bar{\Sigma}$ , then  $A_1 - A_2 \in \bar{\Sigma}$ .

We also expect  $d(\sum A_i, \sum B_i) \leq \sum d(A_i, B_i)$ . The corresponding set theory is  $(\cup A_i) \Delta (\cup B_i) \subset \cup (A_i \Delta B_i)$  and  $(\cap A_i) \Delta (\cap B_i) \subset \cup (A_i \Delta B_i)$ . This implies that if  $A_i \in \bar{\Sigma}$ , then  $\cup A_i, \cap A_i \in \bar{\Sigma}$ .

Next we show that  $\bar{\mu}$  is well defined. Suppose we have

$$A \Delta B_1 \subset C_1, \quad A \Delta B_2 \subset C_2, \quad B_1, B_2, C_1, C_2 \in \Sigma, \quad \mu(C_1) = \mu(C_2) = 0.$$

By (analogue of triangle inequality  $d(B_1, B_2) \leq d(A, B_1) + d(A, B_2)$ )

$$B_1 \Delta B_2 \subset (A \Delta B_1) \cup (A \Delta B_2) \subset C_1 \cup C_2,$$

the assumptions imply  $\mu(B_1) = \mu(B_2)$ .

Now we show that  $\bar{\mu}$  is countably additive. Suppose  $A_i \in \bar{\Sigma}$  are disjoint, and  $A_i$  is almost the same as  $B_i \in \Sigma$ . Then  $B_i \cap B_j$  is almost the same as  $A_i \cap A_j$ , so that  $\mu(B_i \cap B_j) = \bar{\mu}(A_i \cap A_j) = 0$ . The countable additivity of  $\bar{\mu}$  then follows from the fourth property in Proposition 9.4.6.

Finally, we show that  $(X, \bar{\Sigma}, \bar{\mu})$  is complete. Suppose  $A' \subset A \in \bar{\Sigma}$  and  $\bar{\mu}(A) = 0$ . Then we have  $A \Delta B \subset C$  with  $B, C \in \Sigma$  satisfying  $\mu(B) = \bar{\mu}(A) = 0$  and  $\mu(C) = 0$ . This implies that  $B' \Delta \emptyset = B' \subset B \subset A \cup C$ . Since  $A \cup C \in \Sigma$  has measure zero, we conclude that  $B' \in \bar{\Sigma}$ .  $\square$

**Exercise 9.54.** Prove that the completion  $\sigma$ -algebra  $\bar{\Sigma}$  consists of subsets of the form  $A \cup B$ , with  $A \in \Sigma$  and  $B$  contained in a measurable subset of measure zero.

**Exercise 9.55.** Suppose  $(X, \Sigma, \mu)$  is a measure space. For any subset  $A \subset X$ , let

$$\mu^*(A) = \inf\{\mu(C) : A \subset C, C \in \Sigma\}.$$

1. Prove that  $\mu^*$  is an outer measure, and  $\mu^*(A) = \mu(A)$  for  $A \in \Sigma$ .
2. Prove that for any subset  $A$ , there is  $B \in \Sigma$  containing  $A$ , such that  $\mu(B) = \mu^*(A)$ .
3. Prove that any subset in  $\Sigma$  is  $\mu^*$ -measurable.
4. Prove that if  $\mu^*(A) < \infty$ , then for the subset  $B$  in the second part,  $B - A$  is contained in a subset in  $\Sigma$  of measure 0.
5. Prove that if  $(X, \Sigma, \mu)$  is  $\sigma$ -finite, then  $\mu^*$  induces the completion of the measure space  $(X, \Sigma, \mu)$ .
6. Show that  $\sigma$ -finiteness is necessary for the fourth part, by considering  $\Sigma = \{\emptyset, X\}$  and  $\mu(X) = \infty$ .

## 9.5 Additional Exercise

### How Many Subsets are Contained in a $\sigma$ -Algebra

**Exercise 9.56.** Suppose  $\Sigma$  is a  $\sigma$ -algebra on  $X$  that contains infinitely many subsets.

1. Suppose  $A \in \Sigma$  and  $A$  contains infinitely many subsets in  $\Sigma$ . Prove that there is a subset  $B \subset A$ , such that  $B \in \Sigma$ ,  $B \neq A$  and  $B$  contains infinitely many subsets in  $\Sigma$ .
2. Prove that there is a strictly decreasing sequence of subsets in  $\Sigma$ . This implies that  $X = \sqcup_{i=1}^{\infty} A_i$  for some nonempty and disjoint  $A_i \in \Sigma$ .
3. Prove that there are uncountably infinitely many subsets in  $\Sigma$ .

### Saturated Measure

A subset  $A$  is *locally measurable* if  $A \cap B$  is measurable for any measurable  $B$  with finite measure. A measure space is *saturated* if any locally measurable subset is measurable. Let  $\Sigma'$  be the collection of locally measurable subsets in a measure space  $(X, \Sigma, \mu)$ .

**Exercise 9.57.** Prove that  $\sigma$ -finite measure spaces are saturated.



Exercise 9.58. Suppose an outer measure has the property that if  $\mu^*(A) < +\infty$ , then there is a  $\mu^*$ -measurable  $B$  containing  $A$ , such that  $\mu^*(B) < +\infty$ . Prove that  $\mu^*$  induces a saturated measure.

Exercise 9.59. Prove that  $\Sigma'$  is a  $\sigma$ -algebra.

Exercise 9.60. Let  $\mu'$  be the extension of  $\mu$  to  $\Sigma'$  by taking  $\mu' = +\infty$  on  $\Sigma' - \Sigma$ . Prove that  $\mu'$  is a saturated measure.

Exercise 9.61. Prove that  $\mu''(A) = \sup\{\mu(B) : A \supset B \in \Sigma, \mu(B) < +\infty\}$  is the smallest extension of  $\mu$  to a saturated measure on  $\Sigma'$ . Moreover, construct an example for which  $\mu''$  is different from  $\mu'$  in Exercise 9.60.



## Chapter 10

# Lebesgue Integration

## 10.1 Integration in Bounded Case

### Riemann Sum for Lebesgue Integration

Let  $f$  be a bounded function on  $X$ . A *measurable partition* of  $X$  is a finite disjoint union  $X = \sqcup X_i$  of measurable subsets  $X_i$ . For a choice of sample points  $x_i^* \in X_i$ , we may define the “Riemann sum”

$$S(\sqcup X_i, f) = \sum f(x_i^*)\mu(X_i).$$

Like the Riemann integral, the Lebesgue integrability of  $f$  should be the convergence of the sum as the partition gets finer, and the Lebesgue integral should be the limit value.

What do we mean by the partitions becoming finer? Recall that the size of partitions in the definition of Riemann integral is measured by the “diameter” of the pieces. In a general measure space, however, there is no diameter.

How about using  $\max \mu(X_i)$  as the size of a measurable partition  $\sqcup X_i$ ? The first problem is that some measure spaces may not have subsets of small measure (the counting measure in Example 9.4.4, for example). A more serious problem is that the pieces  $X_i$  are allowed to “scatter” all over  $X$  while their measures can still be kept small. For example, consider the presumably integrable function

$$f(x) = \begin{cases} 0, & \text{if } x \in (-1, 0], \\ 1, & \text{if } x \in (0, 1]. \end{cases}$$

The partition

$$(-1, 1] = \cup_{-n \leq i \leq n-1} X_i, \quad X_i = \left( \frac{i}{n}, \frac{i+1}{n} \right) \cup \left\{ -\frac{i}{n} \right\}$$

has the property that any  $X_i$  contains points in  $(-1, 0]$  as well as points in  $(0, 1]$ . Therefore we may choose all  $x_i^* \in (-1, 0]$  to get  $S(\sqcup X_i, f) = 0$ , or all  $x_i^* \in (0, 1]$  to get  $S(\sqcup X_i, f) = 1$ . So the Riemann sum diverges, while the proposed size  $\max \mu(X_i) = \frac{1}{n}$  converges to 0.

So we must take the broader view that a limit does not always have to be taken when  $n \rightarrow \infty$  or  $\delta \rightarrow 0$ . Define a partition  $\sqcup Y_j$  to be a *refinement* of  $\sqcup X_i$  if any  $Y_j$  is contained in some  $X_i$ , and denote the refinement by  $\sqcup Y_j \geq \sqcup X_i$ . The refinement is a *partial order* among all partitions of  $X$  because of the following properties.

- $\sqcup Y_j \geq \sqcup X_i$  and  $\sqcup X_i \geq \sqcup Y_j \implies \sqcup Y_j = \sqcup X_i$ .
- $\sqcup Z_k \geq \sqcup Y_j$  and  $\sqcup Y_j \geq \sqcup X_i \implies \sqcup Z_k \geq \sqcup X_i$ .

Therefore the collection of all partitions  $\sqcup X_i$  form a directed set (see Exercise 2.82 through 2.88). For a fixed  $f$ , we may regard  $S$  as a function on this directed set, and the liminf of such  $S$  can be defined (see Exercises 2.89 through 2.100).

**Definition 10.1.1.** A bounded function  $f$  on a measure space  $(X, \Sigma, \mu)$  with  $\mu(X) < +\infty$  is *Lebesgue integrable*, with  $I = \int_X f d\mu$  as the *Lebesgue integral*, if for any  $\epsilon > 0$ , there is a measurable partition  $\sqcup X_i$ , such that

$$\sqcup Y_j \text{ is a refinement of } \sqcup X_i \implies |S(\sqcup Y_j, f) - I| < \epsilon.$$

The integral of a function  $f(x)$  on an interval  $\langle a, b \rangle$  with respect to the usual Lebesgue measure on  $\mathbb{R}$  is also denoted by  $\int_a^b f(x)dx$ . The notation is the same as the Riemann integral because the values of the Lebesgue and Riemann integrals are expected to be the same.

**Example 10.1.1.** Back to the example above, we take  $\sqcup X_i = (-1, 0] \cup (0, 1]$ . If  $\sqcup Y_j$  is a refinement of  $\sqcup X_i$ , then each  $Y_j$  is contained completely in either  $(-1, 0]$  or  $(0, 1]$ . Therefore

$$\begin{aligned} S(\sqcup Y_j, f) &= \sum_{Y_j \subset (-1, 0]} f(y_j^*)\mu(Y_j) + \sum_{Y_j \subset (0, 1]} f(y_j^*)\mu(Y_j) \\ &= \sum_{Y_j \subset (-1, 0]} 0 \cdot \mu(Y_j) + \sum_{Y_j \subset (0, 1]} 1 \cdot \mu(Y_j) \\ &= \sum_{Y_j \subset (0, 1]} \mu(Y_j) = \mu(\sqcup_{Y_j \subset (0, 1]} Y_j) = \mu(0, 1] = 1. \end{aligned}$$

We conclude that the Lebesgue integral  $\int_{-1}^1 f(x)dx = 1$ .

The example can be extended to the characteristic function

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A, \end{cases}$$

of a measurable subset  $A$ . If  $\sqcup Y_j$  is a refinement of  $A \sqcup (X - A)$ , then  $S(\sqcup Y_j, \chi_A) = \mu(A)$ , and we conclude that  $\int_X \chi_A d\mu = \mu(A)$ . A special case is that the Dirichlet function, as the characteristic function of the set of rational numbers, is Lebesgue integrable with  $\int_{\mathbb{R}} D(x)dx = 0$ . In contrast, we have shown in Example 4.1.4 that the Dirichlet function is not Riemann integrable.

**Exercise 10.1.** Consider the measure  $\mu$  in Exercise 9.38. Prove that if  $X$  is countable and  $\sum \mu_x$  converges, then any bounded function  $f$  is Lebesgue integrable with respect to  $\mu$ , and  $\int_X f d\mu = \sum \mu_x f(x)$ .

### Criterion for Lebesgue Integrability

The Cauchy criterion for the Lebesgue integrability should be the following (see Exercises 2.101 and 2.102): For any  $\epsilon > 0$ , there is a partition  $\sqcup X_i$ , such that

$$\sqcup Y_j, \sqcup Z_k \text{ refine } \sqcup X_i \implies |S(\sqcup Y_j, f) - S(\sqcup Z_k, f)| < \epsilon.$$

Take both  $\sqcup Y_j$  and  $\sqcup Z_k$  to be  $\sqcup X_i$ , with perhaps different choices of  $x_i^* \in X_i$  for  $\sqcup Y_j$  and  $x_i^{**} \in X_i$  for  $\sqcup Z_k$ , we get

$$S(\sqcup Y_j, f) - S(\sqcup Z_k, f) = \sum (f(x_i^*) - f(x_i^{**}))\mu(X_i).$$

By choosing  $x_i^*$  and  $x_i^{**}$  so that  $f(x_i^*)$  is as close to  $\sup_{X_i} f$  as possible and  $f(x_i^{**})$  is as close to  $\inf_{X_i} f$  as possible, the Cauchy criterion implies that the Riemann sum of the oscillations  $\sum \omega_{X_i}(f)\mu(X_i) \leq \epsilon$ . Like the Riemann integral, this property is also sufficient for the Lebesgue integrability.

**Proposition 10.1.2.** *Suppose  $f$  is a bounded function on a measure space  $(X, \Sigma, \mu)$  with  $\mu(X) < +\infty$ . Then  $f$  is Lebesgue integrable if and only if for any  $\epsilon > 0$ , there is a measurable partition  $\sqcup X_i$ , such that  $\sum \omega_{X_i}(f)\mu(X_i) < \epsilon$ .*

Since the integrability criterion in Propositions 4.1.3 is stronger than the one above, we have the following consequence.

**Corollary 10.1.3.** *Riemann integrable functions on an interval are Lebesgue integrable, and the two integrals have the same value.*

On the other hand, Example 10.1.1 shows that the converse is not true. Theorem 10.4.5 gives a necessary and sufficient condition for the Riemann integrability in terms of the Lebesgue measure.

*Proof.* Note that it is easy to show that the integrability implies the Cauchy criterion. The discussion before the proposition then shows that the necessity is a special case of the Cauchy criterion.

Conversely, suppose  $\sum_i \omega_{X_i}(f)\mu(X_i) < \epsilon$  for some measurable partition  $\sqcup X_i$ . If  $\sqcup Y_j$  refines  $\sqcup X_i$ , then  $\mu(X_i) = \sum_{Y_j \subset X_i} \mu(Y_j)$ , and

$$\begin{aligned} |S(\sqcup Y_j, f) - S(\sqcup X_i, f)| &= \left| \sum_j f(y_j^*)\mu(Y_j) - \sum_i f(x_i^*)\mu(X_i) \right| \\ &= \left| \sum_i \left( \sum_{Y_j \subset X_i} f(y_j^*)\mu(Y_j) - f(x_i^*) \sum_{Y_j \subset X_i} \mu(Y_j) \right) \right| \\ &\leq \sum_i \sum_{Y_j \subset X_i} |f(y_j^*) - f(x_i^*)|\mu(Y_j) \\ &\leq \sum_i \sum_{Y_j \subset X_i} \omega_{X_i}(f)\mu(Y_j) = \sum_i \omega_{X_i}(f)\mu(X_i) < \epsilon. \end{aligned}$$

Now let  $\epsilon_n > 0$  converge to 0 and let  $P_n = \sqcup X_i^{(n)}$  be a sequence of partitions, such that the corresponding  $\sum \omega_{X_i^{(n)}}(f)\mu(X_i) < \epsilon_n$ . By what we just proved, if  $P$  refines  $P_n$ , then  $|S(P, f) - S(P_n, f)| < \epsilon_n$ . For any  $m, n$ , by taking intersections,

we can find a partition  $P$  that refine both  $P_m$  and  $P_n$ . Then

$$|S(P_m, f) - S(P_n, f)| \leq |S(P, f) - S(P_m, f)| + |S(P, f) - S(P_n, f)| < \epsilon_m + \epsilon_n.$$

This implies that  $S(P_n, f)$  is a Cauchy sequence of numbers and must converge to a limit  $I$ . Then for any partition  $P$  refining  $P_n$ , we have

$$|S(P, f) - I| \leq |S(P, f) - S(P_n, f)| + |S(P_n, f) - I| < \epsilon_n + |S(P_n, f) - I|.$$

This implies that  $f$  is integrable, with  $I$  as the Lebesgue integral.  $\square$

**Exercise 10.2.** Prove that if  $\sqcup Y_j$  refines  $\sqcup X_i$ , then  $\sum \omega_{Y_j}(f)\mu(Y_j) \leq \sum \omega_{X_i}(f)\mu(X_i)$ .

**Exercise 10.3.** Let  $(X, \Sigma, \mu)$  be a measure space. By Exercise 9.41, for a measurable subset  $A \in \Sigma$ , we have the restriction measure space  $(A, \Sigma_A, \mu_A)$ . For any function  $f$  on  $X$ , prove that the restriction  $f|_A$  is Lebesgue integrable on  $(A, \Sigma_A, \mu_A)$  if and only if  $f\chi_A$  is Lebesgue integrable on  $(X, \Sigma, \mu)$ . Moreover, we have  $\int_A f|_A d\mu_A = \int_X f\chi_A d\mu$ .

## Property of Lebesgue Integral

**Proposition 10.1.4.** *The bounded Lebesgue integrable functions on a measure space  $(X, \Sigma, \mu)$  with  $\mu(X) < +\infty$  have the following properties.*

1. If  $f = g$  almost everywhere, then  $f$  is Lebesgue integrable if and only if  $g$  is Lebesgue integrable, and  $\int_X f d\mu = \int_X g d\mu$ .
2. If  $f$  and  $g$  are Lebesgue integrable, then  $f + g$  and  $cf$  are Lebesgue integrable, and  $\int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu$ ,  $\int_X cf d\mu = c \int_X f d\mu$ .
3. If  $f$  and  $g$  are Lebesgue integrable and  $f \geq g$  almost everywhere, then  $\int_X f d\mu \geq \int_X g d\mu$ . Moreover, if  $\int_X f d\mu = \int_X g d\mu$ , then  $f = g$  almost everywhere.
4. If  $A$  and  $B$  are measurable subsets, then  $f$  is Lebesgue integrable on  $A \cup B$  if and only if it is Lebesgue integrable on  $A$  and on  $B$ . Moreover, we have  $\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu - \int_{A \cap B} f d\mu$ .

The fourth property uses the Lebesgue integral on Lebesgue integral on measurable subsets. The concept is given by Exercise 10.3.

*Proof.* Suppose  $f = g$  away from a subset  $A$  of measure 0. If  $\sqcup X_i$  refines  $A \sqcup (X - A)$ ,

then  $\mu(X_i) = 0$  for those  $X_i \subset A$  and  $f(x_i^*) = g(x_i^*)$  for those  $X_i \subset X - A$ . Therefore

$$\begin{aligned} S(\sqcup X_i, f) &= \sum_{X_i \subset A} f(x_i^*)\mu(X_i) + \sum_{X_i \subset X-A} f(x_i^*)\mu(X_i) \\ &= \sum_{X_i \subset A} f(x_i^*) \cdot 0 + \sum_{X_i \subset X-A} f(x_i^*) \cdot \mu(X_i) \\ &= \sum_{X_i \subset A} g(x_i^*) \cdot 0 + \sum_{X_i \subset X-A} g(x_i^*) \cdot \mu(X_i) = S(\sqcup X_i, g). \end{aligned}$$

This implies that  $f$  is Lebesgue integrable if and only if  $g$  is Lebesgue integrable, and the two integrals are the same. This proves the first property.

Suppose  $f$  and  $g$  are Lebesgue integrable. Then for any  $\epsilon > 0$ , there are partitions  $\sqcup X_i$  and  $\sqcup X'_j$ , such that

$$\begin{aligned} \sqcup Y_k \text{ is a refinement of } \sqcup X_i &\implies |S(\sqcup Y_k, f) - I| < \epsilon, \\ \sqcup Y_k \text{ is a refinement of } \sqcup X'_j &\implies |S(\sqcup Y_k, g) - J| < \epsilon. \end{aligned}$$

Since any refinement  $\sqcup Y_k$  of the partition  $\sqcup_{ij}(X_i \cap X'_j)$  is a refinement of  $\sqcup X_i$  and of  $\sqcup X'_j$ , we get

$$\begin{aligned} |S(\sqcup Y_k, f + g) - I - J| &= |S(\sqcup Y_k, f) + S(\sqcup Y_k, g) - I - J| \\ &\leq |S(\sqcup Y_k, f) - I| + |S(\sqcup Y_k, g) - J| < 2\epsilon. \end{aligned}$$

Here for the equality, the sample points  $x_i^*$  have been chosen for  $f + g$ , and then the same sample points are used for  $f$  and  $g$ . This proves the addition part of the second property. The proof of the scalar multiplication part is similar.

For the third property, we assume  $f \geq g$  almost everywhere. By the first property, we may modify some values of  $f$  and  $g$ , such that  $f \geq g$  everywhere, without affecting the discussion about the property. Then for the same choice of sample points, we have  $S(\sqcup X_i, f) \geq S(\sqcup X_i, g)$ . This further implies  $\int_X f d\mu \geq$

$$\int_X g d\mu.$$

Next we further assume  $\int_X f d\mu = \int_X g d\mu$ . By the second property, the function  $h = f - g \geq 0$  satisfies  $\int_X h d\mu = 0$ . By the definition of  $\int_X h d\mu = 0$ , for any  $\epsilon > 0$ , there is a partition  $\sqcup X_i$ , such that

$$S(\sqcup X_i, h) = \sum h(x_i^*)\mu(X_i) < \epsilon^2.$$

By choosing  $x_i^* \in X_i$  so that  $h(x_i^*)$  is as large as possible, we get

$$\sum \left( \sup_{X_i} h \right) \mu(X_i) \leq \epsilon^2.$$

Let

$$Y_\epsilon = \sqcup \{X_i : h(x) \geq \epsilon \text{ for some } x \in X_i\} = \sqcup \{X_i : \sup_{X_i} h \geq \epsilon\}.$$



Then  $h < \epsilon$  on  $X - Y_\epsilon$ , and by  $h \geq 0$ , we have

$$\epsilon^2 \geq \sum \left( \sup_{X_i} h \right) \mu(X_i) \geq \sum_{\sup_{X_i} h > \epsilon} \epsilon \mu(X_i) = \epsilon \mu(Y_\epsilon).$$

This implies  $\mu(Y_\epsilon) \leq \epsilon$ .

We apply the estimation to a decreasing sequence  $\epsilon_n \rightarrow 0$  and get a corresponding sequence  $Y_{\epsilon_n}$ . Since  $\epsilon_n$  is decreasing, we actually have

$$0 \leq h \leq \epsilon_n \text{ on } \cup_{k \geq n} (X - Y_{\epsilon_k}) = X - Z_n, \quad Z_n = \cap_{k \geq n} Y_{\epsilon_k}.$$

For any  $k \geq n$ , we have  $Z_n \subset Y_{\epsilon_k}$ , so that  $\mu(Z_n) \leq \mu(Y_{\epsilon_k}) \leq \epsilon_k$ . Taking  $k \rightarrow \infty$ , we get  $\mu(Z_n) = 0$ . Then we get

$$h = 0 \text{ on } \cap_n (X - Z_n) = X - \cup_n Z_n, \quad \mu(\cup_n Z_n) = 0.$$

This completes the proof of the third property.

For the fourth property, we consider any partition  $\sqcup X_i$  of  $A \cup B$  that refines  $(A - B) \sqcup (B - A) \sqcup (A \cap B)$ . The integrability can make use of the criterion in Proposition 10.1.2 and

$$\sum_{X_i \subset A} \omega_{X_i}(f) \mu(X_i) \leq \sum_{X_i \subset A \cup B} \omega_{X_i}(f) \mu(X_i)$$

in one direction, and

$$\sum_{X_i \subset A \cup B} \omega_{X_i}(f) \mu(X_i) \leq \sum_{X_i \subset A} \omega_{X_i}(f) \mu(X_i) + \sum_{X_i \subset B} \omega_{X_i}(f) \mu(X_i)$$

in the other direction. Moreover, by taking  $x_i^*$  to be the same, we have

$$S(\sqcup X_i, f) = S(\sqcup_{X_i \subset A-B} X_i, f) + S(\sqcup_{X_i \subset B-A} X_i, f) - S(\sqcup_{X_i \subset A \cap B} X_i, f).$$

This leads to the equality. □

## 10.2 Measurable Function

The integrability criterion in Proposition 10.1.2 says that for a function to be Lebesgue integrable, we need to find partitions so that the oscillations of the function on the pieces in the partitions are *mostly* small. In fact, it is always possible to construct partitions such that the oscillations are *uniformly* small.

For any subset  $A \subset X$ , we have

$$\begin{aligned} \omega_A(f) \leq \epsilon &\iff f(A) \subset [c, c + \epsilon] \text{ for some } c \\ &\iff A \subset f^{-1}[c, d] \text{ for some } c < d \text{ satisfying } d - c \leq \epsilon. \end{aligned}$$

The observation motivates the following construction. For the bounded function  $f$ , we have  $a < f(x) < b$  for some constants  $a$  and  $b$ . Take a partition

$$\Pi: a = c_0 < c_1 < \cdots < c_k = b$$

of  $[a, b]$  satisfying

$$\|\Pi\| = \max_{1 \leq i \leq k} (c_i - c_{i-1}) < \epsilon.$$

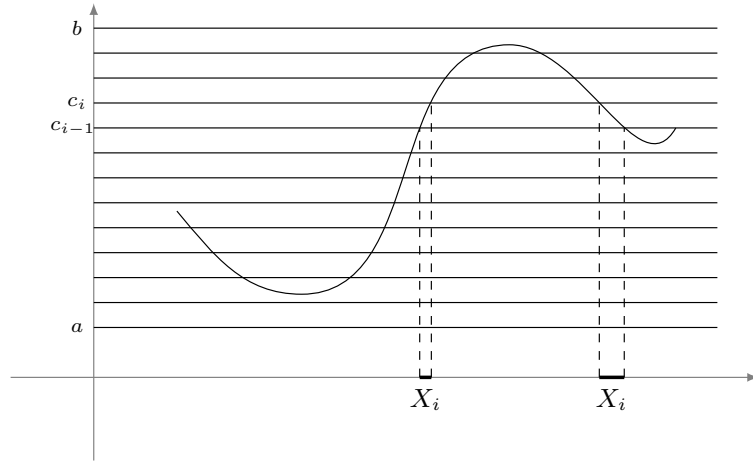
Then the subsets

$$X_i = f^{-1}(c_{i-1}, c_i] = \{x : c_{i-1} < f(x) \leq c_i\}$$

satisfy

$$X = \sqcup X_i, \quad \omega_{X_i}(f) \leq c_i - c_{i-1} < \epsilon.$$

Note that all the oscillations are uniformly less than  $\epsilon$ .



**Figure 10.2.1.** The partition satisfies  $\omega_{X_i}(f) \leq c_i - c_{i-1}$ .

## Measurability of Function

For the construction that produces uniformly small oscillations to work, we need to assume that the pieces  $X_i$  are measurable. In other words,  $f$  should satisfy the following condition.

**Definition 10.2.1.** Let  $\Sigma$  be a  $\sigma$ -algebra on a set  $X$ . A function  $f$  on  $X$  is *measurable* with respect to  $\Sigma$  if for any  $a < b$ , the subset

$$f^{-1}(a, b] = \{x \in X : a < f(x) \leq b\} \in \Sigma$$

is measurable.

**Example 10.2.1.** The characteristic function  $\chi_A$  of a subset  $A$  has the preimages

$$\chi_A^{-1}(a, b] = \begin{cases} X, & \text{if } a < 0 < 1 \leq b, \\ A, & \text{if } 0 \leq a < 1 \leq b, \\ X - A, & \text{if } a < 0 \leq b < 1, \\ \emptyset, & \text{other.} \end{cases}$$

Therefore the function is measurable if and only if  $A \in \Sigma$ .

**Example 10.2.2.** If  $f(x)$  is a monotone function on  $\mathbb{R}$ , then  $f^{-1}(a, b]$  is always an interval. Therefore monotone functions are *Borel measurable* as well as *Lebesgue measurable*.

**Exercise 10.4.** Suppose  $g: X \rightarrow Y$  is a map and  $\Sigma$  is a  $\sigma$ -algebra on  $Y$ . Prove that if  $f$  is a measurable function on  $(Y, \Sigma)$ , then  $f \circ g$  is a measurable function on the preimage  $\sigma$ -algebra  $(X, g^{-1}(\Sigma))$  (see Exercise 9.37).

**Exercise 10.5.** Suppose  $g: X \rightarrow Y$  is a map and  $\Sigma$  is a  $\sigma$ -algebra on  $X$ . Prove that if  $f \circ g$  is a measurable function on  $(X, \Sigma)$ , then  $f$  is a measurable function on the push forward  $\sigma$ -algebra  $(Y, g_*(\Sigma))$  (see Exercise 9.36).

The definition of measurability makes use of the concept of preimage. As shown by the discussion before Proposition 6.4.8, preimage has very nice properties. This leads to many equivalent formulations of the measurability.

Let  $\epsilon_n > 0$  converge to 0. Then we have

$$(a, b) = \cup(a, b - \epsilon_n], \quad (a, b] = \cap(a, b + \epsilon_n).$$

This implies

$$f^{-1}(a, b) = \cup f^{-1}(a, b - \epsilon_n], \quad f^{-1}(a, b] = \cap f^{-1}(a, b + \epsilon_n).$$

Therefore  $f$  is measurable if and only if

$$f^{-1}(a, b) = \{x: a < f(x) < b\} \in \Sigma$$

for any open interval  $(a, b)$ . For example, if  $f$  is a continuous function on  $\mathbb{R}$ , then  $f^{-1}(a, b)$  is an open subset. Therefore continuous functions on  $\mathbb{R}$  are Borel measurable as well as Lebesgue measurable.

By

$$\begin{aligned} f^{-1}(a, b] &= \{x: a < f(x) \leq b\} \\ &= \{x: f(x) > a\} - \{x: f(x) > b\} \\ &= f^{-1}(a, +\infty) - f^{-1}(b, +\infty), \end{aligned}$$

we also see that  $f$  is measurable if and only if

$$f^{-1}(a, +\infty) = \{x: f(x) > a\}$$

is measurable for any  $a$ . Exercise 10.8 contains more equivalent conditions for a function to be measurable.

**Proposition 10.2.2.** *The measurable functions have the following properties.*

1. If  $f$  is measurable and  $g$  is continuous, then  $g \circ f$  is measurable.
2. The arithmetic combinations of measurable functions are measurable.
3. If  $f_n$  is a sequence of measurable functions, then  $\sup f_n$ ,  $\inf f_n$ ,  $\overline{\lim} f_n$ ,  $\underline{\lim} f_n$  are measurable.

4. If  $X = \cup X_i$  is a countable union and  $X_i \in \Sigma$ , then  $f$  is measurable if and only if the restrictions  $f|_{X_i}$  are measurable.

*Proof.* We have  $(g \circ f)^{-1}(a, b) = f^{-1}(g^{-1}(a, b))$ . By Proposition 6.4.8, the continuity of  $g$  implies that  $g^{-1}(a, b)$  is open. By Theorem 6.4.6, we have  $g^{-1}(a, b) = \sqcup (c_i, d_i)$  and then

$$(g \circ f)^{-1}(a, b) = f^{-1}(\sqcup (c_i, d_i)) = \sqcup f^{-1}(c_i, d_i).$$

Since  $f$  is measurable, we have  $f^{-1}(c_i, d_i) \in \Sigma$ , and the countable union is also in  $\Sigma$ .

The measurability of the addition follows from

$$\begin{aligned} (f + g)^{-1}(a, +\infty) &= \{x: f(x) + g(x) > a\} \\ &= \{x: f(x) > r, g(x) > a - r \text{ for some } r \in \mathbb{Q}\} \\ &= \cup_{r \in \mathbb{Q}} (f^{-1}(r, +\infty) \cap g^{-1}(a - r, +\infty)). \end{aligned}$$

Similar argument can be applied to other arithmetic combinations.

The measurability of the supremum follows from

$$\begin{aligned} (\sup f_n(x))^{-1}(a, +\infty) &= \{x: \sup f_n(x) > a\} \\ &= \{x: f_n(x) > a \text{ for some } n\} \\ &= \cup \{x: f_n(x) > a\} \\ &= \cup f_n^{-1}(a, +\infty). \end{aligned}$$

The measurability of the infimum follows from the arithmetic property, the measurability of the supremum, and

$$\inf f_n = -\sup(-f_n).$$

The measurability of the upper limit then follows from

$$\overline{\lim} f_n = \inf_n (\sup \{f_n, f_{n+1}, f_{n+2}, \dots\}).$$

The last property follows from

$$f^{-1}(U) = \cup (f|_{X_i})^{-1}(U), \quad (f|_{X_i})^{-1}(U) = X_i \cap f^{-1}(U). \quad \square$$

**Example 10.2.3 (Semicontinuous Function).** We know continuous functions are Borel measurable. More generally, a real function  $f(x)$  is *lower (upper) semicontinuous* if for any  $x_0$  and  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $|x - x_0| < \delta$  implies  $f(x) > f(x_0) - \epsilon$  ( $f(x) < f(x_0) + \epsilon$ ). Clearly, a function is continuous if and only if it is upper and lower semicontinuous.

The lower semicontinuity at  $x_0$  can be rephrased as that, if  $f(x_0) > a$ , then  $f(x) > a$  for  $x$  sufficiently near  $a$ . This means exactly that  $f^{-1}(a, +\infty)$  is open for any  $a$ . In particular, lower (and upper) semicontinuous functions are Borel measurable.

If  $f_i$  are lower semicontinuous, then  $(\sup f_i(x))^{-1}(a, +\infty) = \cup f_i^{-1}(a, +\infty)$  is still open, so that  $\sup f_i(x)$  is still lower semicontinuous. Note that the number of functions  $f_i$  here does not have to be countable. In particular, the supremum of a (not necessarily

countable) family of continuous functions is lower semicontinuous and Borel measurable. In contrast, we only know that, in general, the supremum of *countably* many measurable functions is measurable.

Now consider a family  $f_t(x)$  of lower semicontinuous functions, with  $t \in (0, \epsilon)$ . We have

$$\overline{\lim}_{t \rightarrow 0^+} f_t = \inf_{\delta \in (0, \epsilon)} \left( \sup_{t \in (0, \delta)} f_t \right).$$

Since  $\sup_{t \in (0, \delta)} f_t$  is increasing in  $\delta$ , for a sequence  $\delta_n > 0$  converging to 0, we have

$$\overline{\lim}_{t \rightarrow 0^+} f_t = \inf_n \left( \sup_{t \in (0, \delta_n)} f_t \right).$$

Now  $\sup_{t \in (0, \delta_n)} f_t$  is lower semicontinuous and Borel measurable. Then the infimum  $\overline{\lim}_{t \rightarrow 0^+} f_t$  of countably many Borel measurable functions is still Borel measurable. In particular, if  $f_t(x)$  is continuous for each  $t \in (0, \epsilon)$ , then  $\overline{\lim}_{t \rightarrow 0^+} f_t$  is Borel measurable.

**Exercise 10.6.** Let  $\Sigma$  be a  $\sigma$ -algebra on  $X$ . Let  $f$  be a function on  $X$ . Prove that the following are equivalent.

1.  $f^{-1}(a, b) = \{x: a < f(x) < b\} \in \Sigma$  for any  $a < b$ .
2.  $f^{-1}[a, b] = \{x: a \leq f(x) \leq b\} \in \Sigma$  for any  $a < b$ .
3.  $f^{-1}(a, b] = \{x: a < f(x) \leq b\} \in \Sigma$  for any  $a < b$ .
4.  $f^{-1}(a, b) = \{x: a < f(x) < b\} \in \Sigma$  for any  $a < b$ ,  $a, b \in \mathbb{Q}$ .
5.  $f^{-1}(a, b) = \{x: a < f(x) < b\} \in \Sigma$  for any  $a, b$  satisfying  $0 < b - a < 1$ .
6.  $f^{-1}(a, a + 1) = \{x: a < f(x) < a + 1\} \in \Sigma$  for any  $a$ .
7.  $f^{-1}[a, +\infty) = \{x: f(x) \geq a\} \in \Sigma$  for any  $a$ .
8.  $f^{-1}(-\infty, a) = \{x: f(x) < a\} \in \Sigma$  for any  $a$ .
9.  $f^{-1}(-\infty, a] = \{x: f(x) \leq a\} \in \Sigma$  for any  $a$ .
10.  $f^{-1}(U) \in \Sigma$  for any open  $U$ .
11.  $f^{-1}(C) \in \Sigma$  for any closed  $C$ .
12.  $f^{-1}(K) \in \Sigma$  for any compact  $K$ .

**Exercise 10.7.** By taking  $Y = \mathbb{R}$  in Exercise 9.36, prove that a function  $f$  is measurable if and only if  $f^{-1}(B)$  is measurable for any Borel set  $B$ .

**Exercise 10.8.** Prove that if  $f$  is measurable and  $\mu(X) < +\infty$ , then  $\lim_{n \rightarrow \infty} \mu(f^{-1}[n, +\infty)) = \lim_{n \rightarrow \infty} \mu(f^{-1}(-\infty, -n]) = 0$ . This means that for any  $\epsilon > 0$ , there is a measurable subset  $A$ , such that  $f$  is bounded on  $A$ , and  $\mu(X - A) < \epsilon$ .

**Exercise 10.9.** Suppose  $f$  and  $g$  are functions on a complete measure space, such that  $f = g$  almost everywhere. Prove that  $f$  is measurable if and only if  $g$  is.

**Exercise 10.10.** Theorem 11.4.6 will give us subsets that are not Lebesgue measurable. Use this to construct a family of lower semicontinuous functions, such that the supremum and infimum are not Lebesgue measurable.

Exercise 10.11. Prove properties of semicontinuous functions.

1. If  $f$  and  $g$  are lower semicontinuous and  $c > 0$ , then  $f + g$  and  $cf$  are lower semicontinuous.
2.  $f$  is upper semicontinuous if and only if  $-f$  is lower semicontinuous.
3. If  $f$  is lower semicontinuous and  $g$  is continuous, then the composition  $f \circ g$  is lower semicontinuous.
4. If  $f$  is lower semicontinuous and  $g$  is increasing and left continuous, then the composition  $g \circ f$  is lower semicontinuous.

Exercise 10.12. Prove that the *lower envelope*

$$f_*(x) = \lim_{\delta \rightarrow 0^+} \inf_{(x-\delta, x+\delta)} f$$

is the biggest lower semicontinuous function no bigger than  $f$ . Similarly, the *upper envelope*

$$f^*(x) = \lim_{\delta \rightarrow 0^+} \sup_{(x-\delta, x+\delta)} f$$

is the smallest upper semicontinuous function no smaller than  $f$ .

## Integrability v.s. Measurability

Now we come back to the Lebesgue integrability in terms of the measurability.

**Theorem 10.2.3.** *Suppose  $f$  is a bounded function on a measure space  $(X, \Sigma, \mu)$  with  $\mu(X) < +\infty$ . Then  $f$  is Lebesgue integrable if and only if it is equal to a measurable function almost everywhere.*

*Proof.* For the sufficiency, by the first property of Proposition 10.1.4, we only need to show that the measurability implies the integrability. Suppose  $a < f < b$ . We take a partition  $\Pi$  of  $[a, b]$  and construct

$$X_i = f^{-1}(c_{i-1}, c_i] = \{x : c_{i-1} < f(x) \leq c_i\},$$

Then  $X = \sqcup X_i$  is a measurable partition,  $\omega_{X_i}(f) \leq c_i - c_{i-1} \leq \|\Pi\|$ , and we have

$$\sum \omega_{X_i}(f) \mu(X_i) \leq \|\Pi\| \sum \mu(X_i) \leq \|\Pi\| \mu(X).$$

Since  $\mu(X)$  is finite, by taking  $\|\Pi\|$  to be as small as we want, the integrability criterion in Proposition 10.1.2 is satisfied, and  $f$  is integrable.

For the necessity, we use the integrability criterion in Proposition 10.1.2. If  $f$  is integrable, then for any  $\epsilon > 0$ , there is a measurable partition  $\sqcup X_i$ , such that  $\sum \omega_{X_i}(f) \mu(X_i) < \epsilon$ . Then the functions (upper and lower approximations of  $f$  with respect to the partition  $\sqcup X_i$ )

$$\phi_\epsilon = \sum \left( \inf_{X_i} f \right) \chi_{X_i}, \quad \psi_\epsilon = \sum \left( \sup_{X_i} f \right) \chi_{X_i},$$

are measurable. We also have  $\phi_\epsilon \leq f \leq \psi_\epsilon$ , and

$$\int_X (\psi_\epsilon - \phi_\epsilon) d\mu = \sum \left( \sup_{X_i} f - \inf_{X_i} f \right) \mu(X_i) = \sum \omega_{X_i}(f) \mu(X_i) < \epsilon.$$

Let  $\epsilon_n > 0$  converge to 0. For each  $\epsilon_n$ , we find the measurable partition and the corresponding upper and lower approximations  $\phi_{\epsilon_n}$  and  $\psi_{\epsilon_n}$ . By Proposition 10.2.2, the functions

$$g = \sup \phi_{\epsilon_n}, \quad h = \inf \psi_{\epsilon_n},$$

are measurable. By  $\phi_{\epsilon_n} \leq f \leq \psi_{\epsilon_n}$ , we also have  $\phi_{\epsilon_n} \leq g \leq f \leq h \leq \psi_{\epsilon_n}$ , so that

$$0 \leq \int_X (h - g) d\mu \leq \int_X (\psi_{\epsilon_n} - \phi_{\epsilon_n}) d\mu \leq \epsilon_n.$$

Since this holds for any  $n$ , we have  $\int_X (h - g) d\mu = 0$ . By the fourth property in Proposition 10.1.4, we have  $g = h$  almost everywhere. This implies that  $f = g$  almost everywhere.  $\square$

**Example 10.2.4.** In Example 10.1.1, we see that the characteristic function  $\chi_A$  of a measurable subset  $A$  is integrable. In general, Theorem 10.2.3 says that the integrability of  $\chi_A$  implies  $\chi_A = g$  almost everywhere, for a measurable  $g$ . Then  $B = g^{-1}(1)$  is measurable. Since both  $A - B = \{x: \chi_A(x) = 1 \neq g(x)\}$  and  $B - A = \{x: \chi_A(x) \neq 1 = g(x)\}$  are contained in the subset where  $\chi_A$  and  $g$  are not equal,  $(A - B) \cup (B - A)$  is contained in a subset of measure 0. This shows that  $A$  is almost the same as the measurable subset  $B$ . This further implies  $\chi_A = \chi_B$  almost everywhere and  $\int_X \chi_A d\mu = \int_X \chi_B d\mu = \mu(B)$  (the second equality is due to Example 10.1.1).

We conclude that  $\chi_A$  is integrable if and only if  $A$  is almost the same as a measurable subset.

**Exercise 10.13.** Suppose  $f$  is Lebesgue integrable. Prove that  $f^{-1}(a, b]$  is almost the same as a measurable subset for any interval  $(a, b]$ .

**Exercise 10.14.** Suppose  $f$  is a bounded and Lebesgue integrable function on a measure space  $(X, \Sigma, \mu)$  with  $\mu(X) < +\infty$ .

1. If  $\int_A f d\mu = 0$  for all  $A \in \Sigma$ , prove that  $f = 0$  almost everywhere.
2. If  $f \geq 0$  and  $\int_X f d\mu = 0$ , prove that  $f = 0$  almost everywhere.

The second part gives an alternative proof for the third property in Proposition 10.1.4.

**Exercise 10.15.** Suppose  $f$  is bounded on  $\mathbb{R}$  and equal to a measurable function almost everywhere. Prove that if  $\int_a^b f dx = 0$  on any bounded interval  $(a, b)$ , then  $f = 0$  almost everywhere.

**Exercise 10.16.** A *random variable* on a measure space (actually on a *probability space*, meaning that the measure of the whole space is 1) is simply a measurable function  $f$ . The *distribution function* of the random variable is

$$\alpha(x) = \mu(f^{-1}(-\infty, x]).$$

Assume  $a < f < b$  and  $\alpha$  does not take  $+\infty$  value.

1. For any partition  $a = c_0 < c_1 < \cdots < c_n = b$ , prove that  $\sum c_{i-1} \Delta \alpha_i \leq \int_X f d\mu \leq \sum c_i \Delta \alpha_i$ .
2. Prove that  $\int_X f d\mu = \int_a^b x d\alpha$ , where the right side is the Riemann-Stieltjes integral.

### 10.3 Integration in Unbounded Case

Let  $(X, \Sigma, \mu)$  be a measure space, possibly with  $\mu(X) = +\infty$ . Let  $f$  be an extended valued function on  $X$ . We define the integration of  $f$  by reducing to the bounded case, by considering the integration of the *truncated function*

$$f_{[a,b]} = \begin{cases} f(x), & \text{if } a \leq f(x) \leq b, \\ b, & \text{if } f(x) > b, \\ a, & \text{if } f(x) < a, \end{cases}$$

on measurable subsets  $A$  with  $\mu(A) < +\infty$ . Then we expect to have the limit

$$\int_X f d\mu = \lim_{\substack{\mu(A) < +\infty, A \rightarrow X \\ a \rightarrow -\infty, b \rightarrow +\infty}} \int_A f_{[a,b]} d\mu.$$

**Definition 10.3.1.** Suppose a (possibly extended valued) function  $f$  is equal to a measurable function almost everywhere. If there is  $I$ , such that for any  $\epsilon > 0$ , there is  $N > 0$  and  $Y \in \Sigma$  with  $\mu(Y) < +\infty$ , such that

$$Y \subset A \in \Sigma, \mu(A) < +\infty, a < -N, b > N \implies \left| \int_A f_{[a,b]} d\mu - I \right| < \epsilon,$$

then  $f$  is *Lebesgue integrable* with  $\int_X f d\mu = I$ .

The assumption about equal to a measurable function almost everywhere is to make sure that the integral  $\int_A f_{[a,b]} d\mu$  makes sense. Therefore the integrability here only means the *convergence* of the limit of “partial integrals”, similar to the convergence of improper Riemann integral.

For  $f \geq 0$ , we are concerned with the convergence of

$$\int_X f d\mu = \lim_{\substack{\mu(A) < +\infty, A \rightarrow X \\ b \rightarrow +\infty}} \int_A f_{[0,b]} d\mu.$$



Since the integral  $\int_A f_{[0,b]} d\mu$  becomes bigger as  $A$  or  $b$  becomes bigger, the definition becomes

$$\int_X f d\mu = \sup_{\mu(A) < +\infty, b > 0} \int_A f_{[0,b]} d\mu. \quad (10.3.1)$$

In particular, the convergence is equivalent to the boundedness of the partial integrals.

**Example 10.3.1.** By the argument in Example 10.2.4, a characteristic function  $\chi_A$  is equal to a measurable function almost everywhere if and only if  $A$  is almost equal to a measurable subset  $B$ . We ask whether the equality in Example 10.1.1

$$\int_X \chi_A d\mu = \mu(B)$$

still holds.

By (10.3.1), we have

$$\begin{aligned} \int_X \chi_A d\mu &= \sup_{C \in \Sigma, \mu(C) < +\infty} \int_C \chi_A d\mu \\ &= \sup_{C \in \Sigma, \mu(C) < +\infty} \mu(C \cap A) \\ &= \sup_{B \supset C \in \Sigma, \mu(C) < +\infty} \mu(C). \end{aligned}$$

If  $\mu(B) < +\infty$ , then we may take  $C = B$ , and the right side is indeed  $\mu(B)$ . If  $\mu(B) = +\infty$ , however, the right side is not necessarily  $+\infty$ . For example, let  $\Sigma$  be all subsets of  $X$ , and let  $\mu(A) = +\infty$  for all nonempty  $A \subset X$ , and  $\mu(\emptyset) = 0$ . Then we can only choose  $C = \emptyset$  in the formula above, and we find that the integration is always 0. The counterexample also shows that Exercise 10.14 is no longer valid for the unbounded case in general.

**Definition 10.3.2.** A measure  $\mu$  is *semifinite* if for any measurable  $A$  with  $\mu(A) = +\infty$ , there is measurable  $B \subset A$ , such that  $0 < \mu(B) < +\infty$ .

**Proposition 10.3.3.** The formula  $\int_X \chi_A d\mu = \mu(A)$  holds for all measurable subsets  $A$  if and only if the measure is semifinite.

*Proof.* As explained in Example 10.3.1, all we need to show is that, if  $\mu(A) = +\infty$ , then  $\alpha = \sup\{\mu(B) : A \supset B \in \Sigma, \mu(B) < +\infty\}$  is  $+\infty$ .

If  $\alpha < +\infty$ , then  $\alpha$  is finite and  $\alpha = \lim \mu(B_n)$  for a sequence  $B_n \subset A$  satisfying  $\mu(B_n) < +\infty$ . By  $B_1 \cup \dots \cup B_n \subset A$ ,  $\mu(B_1 \cup \dots \cup B_n) < +\infty$  and the definition of  $\alpha$ , we have  $\mu(B_n) \leq \mu(B_1 \cup \dots \cup B_n) < \alpha$ . Since  $B_1 \cup \dots \cup B_n$  is an increasing sequence, by the third part of Proposition 9.4.4 and the sandwich rule, we find that  $B = \bigcup B_n \subset A$  satisfies  $\mu(B) = \lim \mu(B_1 \cup \dots \cup B_n) = \alpha$ .

Since  $\mu(A) = +\infty$  and  $\mu(B) = \alpha < +\infty$ , we get  $\mu(A - B) = +\infty$ . By the semifinite property, there is a measurable  $C \subset A - B$ , such that  $0 < \mu(C) < +\infty$ . Then the measure of  $B \cup C \subset A$  satisfies

$$\alpha < \mu(B \cup C) = \mu(B) + \mu(C) = \alpha + \mu(C) < +\infty.$$

This contradicts to the definition of  $\alpha$ . □

Exercise 10.17. Extend Exercise 10.3 to the unbounded case.

Exercise 10.18. Prove that any  $\sigma$ -finite measure is semifinite. Moreover, show that the converse is not true.

Exercise 10.19. Let  $\mu$  be a measure on a  $\sigma$ -algebra  $\Sigma$ . For  $A \in \Sigma$ , let

$$\mu_1(A) = \sup\{\mu(B) : A \supset B \in \Sigma, \mu(B) < +\infty\}.$$

1. Prove that, if  $\mu_1(A) < +\infty$ , then there is  $A \supset B \in \Sigma$ , such that  $\mu(B) = \mu_1(A)$ , and any  $A - B \supset C \in \Sigma$  has  $\mu(C) = 0$  or  $+\infty$ .
2. Prove that  $\mu_1$  is a semifinite measure.

Exercise 10.20. Prove that any measure is a sum  $\mu_1 + \mu_2$ , with  $\mu_1$  semifinite and  $\mu_2$  only taking values 0 and  $+\infty$ . Moreover, prove that the semifinite measure  $\mu_1$  in Exercise 10.19 is the smallest in such decompositions.

To simplify the subsequent discussions, we will assume all functions are measurable.

### Convergence v.s. Boundedness

We split any function into positive and negative parts

$$f = f^+ - f^-, \quad f^+ = \max\{f, 0\}, \quad f^- = -\min\{f, 0\} = \max\{-f, 0\}.$$

Then  $f_{[a,b]} = f_{[0,b]} + f_{[a,0]} = f_{[0,b]}^+ - f_{[0,-a]}^-$  for  $a < 0 < b$ , and

$$\int_A f_{[a,b]} d\mu = \int_A f_{[0,b]}^+ d\mu - \int_A f_{[0,-a]}^- d\mu. \quad (10.3.2)$$

Since  $a$  and  $b$  are independent, we expect that  $f$  is integrable (meaning convergence) if and only if  $f^+$  and  $f^-$  are integrable (see Proposition 10.3.5 below), and we have

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

This reduces the convergence problem to the case of non-negative functions.

**Definition 10.3.4.** A Lebesgue integral  $\int_X f d\mu$  is *bounded* if there is a number  $M$ , such that

$$\left| \int_A f_{[a,b]} d\mu \right| < M \text{ for any } A \in \Sigma \text{ with } \mu(A) < +\infty \text{ and any } a < b.$$

The following says that the convergence of the Lebesgue integral is always absolute. So there is no conditional convergence for the Lebesgue integral.

**Proposition 10.3.5.** *A measurable function is Lebesgue integrable if and only if the integral is bounded. Moreover, the following are equivalent.*

1.  $f$  is Lebesgue integrable.
2.  $f^+$  and  $f^-$  are Lebesgue integrable.
3.  $|f|$  is Lebesgue integrable.

*Proof.* Denote by  $I(f)$  the integrability of  $f$  and by  $B(f)$  the boundedness of  $\int_X f d\mu$ . Then (10.3.1) implies

$$I(f^+) \iff B(f^+), \quad I(f^-) \iff B(f^-), \quad I(|f|) \iff B(|f|).$$

By  $f = f^+ - f^-$  and (10.3.2), we have

$$I(f^+) \text{ and } I(f^-) \implies I(f), \quad B(f^+) \text{ and } B(f^-) \implies B(f).$$

By  $|f| = f^+ + f^-$ ,  $|f^+| \leq |f|$ ,  $|f^-| \leq |f|$ , it is also easy to see that

$$B(f^+) \text{ and } B(f^-) \iff B(|f|).$$

It remains to show that

$$B(f) \implies B(f^+) \text{ and } B(f^-), \quad I(f) \implies B(f^+) \text{ and } B(f^-).$$

Given  $B(f)$ , we have the inequality in Definition 10.3.4. Applying the inequality to  $[0, b]$ , we get

$$\left| \int_A f_{[0,b]}^+ d\mu \right| = \left| \int_A f_{[0,b]} d\mu \right| < M.$$

This proves that  $B(f) \implies B(f^+)$ . The same proof shows that  $B(f) \implies B(f^-)$ .

Given  $I(f)$ , for  $\epsilon = 1$ , there is  $N > 0$  and  $Y \in \Sigma$  with  $\mu(Y) < +\infty$ , such that

$$Y \subset A \in \Sigma, \mu(A) < +\infty, a \leq -N, b \geq N \implies \left| \int_A f_{[a,b]} d\mu - I \right| < 1. \quad (10.3.3)$$

We will argue that this implies the boundedness of  $\int_A f_{[0,b]} d\mu$  for all measurable  $A$  satisfying  $\mu(A) < +\infty$ , and  $A \cap Y = \emptyset$  or  $A \subset Y$ . Then for general measurable  $A$  satisfying  $\mu(A) < +\infty$ , we get  $B(f^+)$  by

$$\int_A f_{[0,b]}^+ d\mu = \int_A f_{[0,b]} d\mu = \int_{A-Y} f_{[0,b]} d\mu + \int_{A \cap Y} f_{[0,b]} d\mu.$$

The argument for  $B(f^-)$  is similar.

Suppose  $A$  satisfies  $\mu(A) < +\infty$  and  $A \cap Y = \emptyset$ . Then  $A^+ = A \cap f^{-1}[0, +\infty)$  has the same property, and for  $a = -N$  and  $b \geq N$ , we have

$$\begin{aligned} \int_A f_{[0,b]} d\mu &= \int_{A^+} f_{[0,b]} d\mu = \left| \int_{A^+} f_{[-N,b]} d\mu \right| = \left| \int_{A^+ \cup Y} f_{[-N,b]} d\mu - \int_Y f_{[-N,b]} d\mu \right| \\ &\leq \left| \int_{A^+ \cup Y} f_{[-N,b]} d\mu - I \right| + \left| \int_Y f_{[-N,b]} d\mu - I \right| < 2. \end{aligned}$$

Here (10.3.3) applied to  $A^+ \cup Y$  and  $Y$  in the last inequality.

Suppose  $A$  satisfies  $\mu(A) < +\infty$  and  $A \subset Y$ . Then for  $a = -N$  and  $b \geq N$ , we have

$$\begin{aligned} \int_A f_{[0,b]} d\mu &\leq \int_Y f_{[0,b]} d\mu = \left| \int_Y f_{[-N,b]} d\mu - \int_Y f_{[-N,0]} d\mu \right| \\ &\leq \left| \int_Y f_{[-N,b]} d\mu - I \right| + |I| + \left| \int_Y f_{[-N,0]} d\mu \right| < 1 + |I| + N\mu(Y). \end{aligned}$$

We note that although the boundedness is only verified to  $b \geq N$ , we actually get the boundedness for  $b \geq 0$  by  $f_{[0,b]} \leq f_{[0,N]}$  in case  $0 \leq b \leq N$ .  $\square$

**Exercise 10.21 (Comparison Test).** Suppose  $f$  and  $g$  are measurable. Prove that if  $|f| \leq g$  for an integrable  $g$ , then  $f$  is integrable.

**Exercise 10.22.** Extend Exercise 10.1 to unbounded case. For the measure  $\mu$  in Exercise 9.38 and any function  $f$ , we have  $\int f d\mu = \sum \mu_x f(x)$ . The function is integrable if and only if  $\sum \mu_x f(x)$  contains only countably many nonzero terms and the sum absolutely converges.

## Property of Lebesgue Integral

**Proposition 10.3.6.** *The (unbounded) Lebesgue integration has the following properties.*

1. If  $f = g$  almost everywhere, then  $f$  is Lebesgue integrable if and only if  $g$  is Lebesgue integrable, and  $\int_X f d\mu = \int_X g d\mu$ .
2. If  $f$  and  $g$  are Lebesgue integrable, then  $f + g$  and  $cf$  are Lebesgue integrable, and  $\int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu$ ,  $\int_X cf d\mu = c \int_X f d\mu$ .
3. If  $f$  and  $g$  are Lebesgue integrable and  $f \geq g$  almost everywhere, then  $\int_X f d\mu \geq \int_X g d\mu$ . Moreover, if  $X$  is semifinite and  $\int_X f d\mu = \int_X g d\mu$ , then  $f = g$  almost everywhere.

4. If  $A$  and  $B$  are measurable subsets, then  $f$  is Lebesgue integrable on  $A \cup B$  if and only if it is Lebesgue integrable on  $A$  and on  $B$ . Moreover, we have
- $$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu - \int_{A \cap B} f d\mu.$$

*Proof.* If  $f = g$  almost everywhere, then  $f_{[a,b]} = g_{[a,b]}$  almost everywhere. By the first property of Proposition 10.1.4, we have  $\int_A f_{[a,b]} d\mu = \int_A g_{[a,b]} d\mu$  for  $\mu(A) < +\infty$ . Then by Definition 10.3.1, we find that  $f$  is integrable if and only if  $g$  is integrable, and their integrals are the same. This proves the first property.

Next we prove the fourth property. By using the boundedness criterion in Proposition 10.3.5, we find that the integrability of  $f$  on measurable  $A$  implies the integrability of  $f$  on measurable subsets of  $A$ . In particular, the integrability on  $A \cup B$  implies the integrability on  $A$  and on  $B$ .

Conversely, if  $C \subset A \cup B$  is measurable with  $\mu(C) < +\infty$ , then we have

$$\int_C f_{[a,b]} d\mu = \int_{C \cap A} f_{[a,b]} d\mu + \int_{C \cap B} f_{[a,b]} d\mu - \int_{C \cap A \cap B} f_{[a,b]} d\mu.$$

If  $f$  is integrable on  $A$  and on  $B$ , then  $f$  is also integrable on the subset  $A \cap B$ . By the boundedness criterion in Proposition 10.3.5 and the equality above, we find that the integrability on  $A$  and on  $B$  implies the integrability on  $A \cup B$ .

For the equality in the fourth property, we start from the definition of the three integrals on the right. For any  $\epsilon > 0$ , there is  $N > 0$  and measurable  $Y \subset A$ ,  $Z \subset B$ ,  $W \subset A \cap B$  with finite measures, such that

$$\begin{aligned} Y \subset C \subset A, \mu(C) < +\infty, a < -N, b > N &\implies \left| \int_C f_{[a,b]} d\mu - I \right| < \epsilon, \\ Z \subset C \subset B, \mu(C) < +\infty, a < -N, b > N &\implies \left| \int_C f_{[a,b]} d\mu - J \right| < \epsilon, \\ W \subset C \subset A \cap B, \mu(C) < +\infty, a < -N, b > N &\implies \left| \int_C f_{[a,b]} d\mu - K \right| < \epsilon, \end{aligned}$$

Now for  $Y \cup Z \cup W \subset C \subset A \cup B$ ,  $\mu(C) < +\infty$  and  $a < -N$ ,  $b > N$ , we apply the implications to  $Y \subset C \cap A \subset A$ ,  $Z \subset C \cap B \subset B$ ,  $W \subset C \cap A \cap B \subset A \cap B$ , and then get

$$\begin{aligned} &\left| \int_C f_{[a,b]} d\mu - I - J - K \right| \\ &\leq \left| \int_{C \cap A} f_{[a,b]} d\mu - I \right| + \left| \int_{C \cap B} f_{[a,b]} d\mu - J \right| + \left| \int_{C \cap A \cap B} f_{[a,b]} d\mu - K \right| < 3\epsilon. \end{aligned}$$

This proves the equality.

Now we turn to the second property. For integrable  $f$  and  $c > 0$ , we have  $(cf)_{[a,b]} = cf_{[c^{-1}a, c^{-1}b]}$ . For integrable  $f$  and  $c < 0$ , we have  $(cf)_{[a,b]} = cf_{[c^{-1}b, c^{-1}a]}$ .

Then we get  $\int_X cf d\mu = c \int_X f d\mu$  by Definition 10.3.1.

For  $f, g \geq 0$  and  $b \geq 0$ , we have

$$f_{[0,b]} + g_{[0,b]} \leq (f + g)_{[0,2b]} \leq f_{[0,2b]} + g_{[0,2b]}.$$

Then either by Definition 10.3.1 or by (10.3.1), we get

$$\int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

This proves the addition formula for non-negative functions.

In general, we consider integrable  $f, g$  and decompose  $X$  into disjoint union of the following six subsets. In each case, we also indicate sum of non-negative functions.

1.  $f \geq 0, g \geq 0$ :  $\text{sum } (f + g) = f + g$ .
2.  $f < 0, g < 0$ :  $\text{sum } -(f + g) = (-f) + (-g)$ .
3.  $f \geq 0, g < 0, f + g \geq 0$ :  $\text{sum } f = (f + g) + (-g)$ .
4.  $f \geq 0, g < 0, f + g < 0$ :  $\text{sum } -g = f + (-(f + g))$ .
5.  $f < 0, g \geq 0, f + g \geq 0$ :  $\text{sum } g = (f + g) + (-f)$ .
6.  $f < 0, g \geq 0, f + g < 0$ :  $\text{sum } -f = g + (-(f + g))$ .

In each case, we may apply the addition formula for non-negative functions, with possible use of the scalar property for  $c = -1$ . For example, on the third subset, we have  $0 \leq f + g < f$ . Then by the boundedness criterion in Proposition 10.3.5, the integrability of  $f$  implies the integrability of  $f + g$ . Then the integrability of non-negative functions  $f + g$  and  $-g$  implies

$$\int f d\mu = \int (f + g) d\mu + \int (-g) d\mu$$

on the third subset. Using the scalar property, this is the same as

$$\int (f + g) d\mu = \int f d\mu + \int g d\mu$$

on the third subset. Then we may use the fourth property that we just proved to add the six equalities together. What we get is the addition formula.

Finally, we prove the third property. If  $f \geq g$  almost everywhere, then  $f_{[a,b]} \geq g_{[a,b]}$  almost everywhere. By the third property of Proposition 10.1.4, we have  $\int_A f_{[a,b]} d\mu \geq \int_A g_{[a,b]} d\mu$  for  $\mu(A) < +\infty$ . Then by Definition 10.3.1, we find that  $\int_X f d\mu \geq \int_X g d\mu$ .

Now we additionally assume that  $\int_X f d\mu = \int_X g d\mu$ . Since  $f$  and  $g$  are integrable, by the second property,  $h = f - g \geq 0$  is also integrable, and  $\int_X h d\mu =$

$\int_X f d\mu - \int_X g d\mu = 0$ . For any  $\epsilon > 0$ , let  $Y_\epsilon = \{x: h(x) \geq \epsilon\}$ . Then  $f \geq \epsilon \chi_{Y_\epsilon}$ . If  $X$  is semifinite, then we have

$$\begin{aligned} 0 &= \int_X h d\mu \geq \int_X \epsilon \chi_{Y_\epsilon} d\mu && \text{(inequality part of third property)} \\ &= \epsilon \int_X \chi_{Y_\epsilon} d\mu && \text{(second property)} \\ &= \epsilon \mu(Y_\epsilon). && \text{(Proposition 10.3.3)} \end{aligned}$$

This implies  $\mu(Y_\epsilon) = 0$  for any  $\epsilon > 0$  and proves that  $h = 0$  almost everywhere.  $\square$

**Exercise 10.23.** Suppose  $f$  is an extended valued measurable function that is Lebesgue integrable on a  $\sigma$ -finite (or semifinite) measure space.

1. Prove that  $\lim_{b \rightarrow +\infty} \mu\{x: |f(x)| \geq b\} = 0$ .
2. Prove that  $\{x: f(x) = \pm\infty\}$  has zero measure.
3. Prove that there is a measurable function  $g$  without taking  $\pm\infty$  value, such that  $f = g$  almost everywhere.

## Infinite Value of Lebesgue Integral

The Lebesgue integral can be naturally extended to have infinity values.

**Definition 10.3.7.** Suppose a (possibly extended valued) function  $f$  is equal to a measurable function almost everywhere. We say the Lebesgue integral of  $f$  diverges to  $+\infty$ , and denote  $\int_X f d\mu = +\infty$ , if for any  $M$ , there is  $N > 0$  and  $Y \in \Sigma$  with  $\mu(Y) < +\infty$ , such that

$$Y \subset A \in \Sigma, \mu(A) < +\infty, a < -N, b > N \implies \int_A f_{[a,b]} d\mu > M.$$

Similarly, we have  $\int_X f d\mu = -\infty$  if

$$Y \subset A \in \Sigma, \mu(A) < +\infty, a < -N, b > N \implies \int_A f_{[a,b]} d\mu < M.$$

We could certainly also define  $\int_X f d\mu = \infty$  by requiring

$$Y \subset A \in \Sigma, \mu(A) < +\infty, a < -N, b > N \implies \left| \int_A f_{[a,b]} d\mu \right| > M.$$

However, such definition is usually not accepted because of two practical concerns. First, the extended arithmetic with unsigned infinity is messy. Second, the unsigned infinity is not considered as an extended value, and the concept of “indefinite integral measure” (see Proposition 10.3.9 and Definition 12.1.1) does not allow unsigned infinity.

**Proposition 10.3.8.**  $\int_X f d\mu = +\infty$  if and only if  $\int_X f^+ d\mu$  is unbounded and  $\int_X f^- d\mu$  is bounded.  $\int_X f d\mu = -\infty$  if and only if  $\int_X f^+ d\mu$  is bounded and  $\int_X f^- d\mu$  is unbounded.

*Proof.* If  $\int_X f^- d\mu$  is bounded, then there is  $M_1$ , such that for all  $A$  with  $\mu(A) < +\infty$  and  $a < 0$ , we have

$$\int_A f_{[a,0]} d\mu = - \int_A f_{[0,-a]}^- d\mu \geq M_1.$$

If we further know that  $\int_A f^+ d\mu$  is unbounded, then for the  $M_1$  obtained above and any  $M$ , there are  $N > 0$  and  $\mu(Y) < +\infty$ , such that

$$Y \subset A, \mu(A) < +\infty, b > N \implies \int_A f_{[0,b]} d\mu = \int_A f_{[0,b]}^+ d\mu > M - M_1.$$

By the second property of Proposition 10.1.4, therefore,  $Y \subset A$ ,  $\mu(A) < +\infty$  and  $b > N$  imply

$$\int_A f_{[a,b]} d\mu = \int_A f_{[a,0]} d\mu + \int_A f_{[0,b]} d\mu > M_1 + (M - M_1) = M.$$

This proves that  $\int_X f d\mu = +\infty$ .

Conversely, suppose  $\int_X f d\mu = +\infty$ . Then for  $M = 0$ , there are  $N > 0$  and  $\mu(Y) < +\infty$ , such that

$$Y \subset A, \mu(A) < +\infty, a \leq -N \implies \int_A f_{[a,N]} d\mu > 0.$$

Let  $A' = \{x \in A : f(x) \leq 0\} \cup Y$ . Then

$$Y \subset A' \subset A, \quad f_{[a,0]} = 0 \text{ on } A - A', \quad f_{[0,N]} = 0 \text{ on } A' - Y,$$

and we get

$$0 < \int_{A'} f_{[a,N]} d\mu = \int_{A'} f_{[a,0]} d\mu + \int_{A'} f_{[0,N]} d\mu = \int_A f_{[a,0]} d\mu + \int_Y f_{[0,N]} d\mu.$$

This further implies

$$\int_A f_{[0,-a]}^- d\mu = - \int_A f_{[a,0]} d\mu < \int_Y f_{[0,N]} d\mu \leq N\mu(Y).$$

Since the right side depends only on  $Y$ , we conclude that  $\int_X f^- d\mu$  is bounded. Then

by Proposition 10.3.5,  $\int_X f d\mu = +\infty$  implies that  $\int_X f^+ d\mu$  cannot be bounded.

Since  $f^+ \geq 0$ , this means  $\int_X f^+ d\mu = +\infty$ . □



**Example 10.3.2.** Example 10.3.1 shows that, if  $\mu$  is semifinite, then  $\mu(A) = +\infty$  implies  $\int_X \chi_A d\mu = +\infty$ . In fact, the property is equivalent to the semifiniteness of  $\mu$ .

**Example 10.3.3.** We show that it is possible for the value of the Lebesgue integral to be infinity but without definite sign.

Let  $X$  be the set of non-negative integers. Let  $\mu\{n\} = a^n$  for a fixed  $a > 2$  (see Exercise 9.38). Let  $f(n) = (-1)^n$ . A subset  $A$  has finite measure if and only if  $A$  is a finite subset. Moreover, we have

$$\left| \int_A f d\mu \right| = \left| \sum_{n \in A} (-1)^n a^n \right| \geq a^{\max A} - \sum_{n=0}^{\max A-1} a^n = a^{\max A} - \frac{a^{\max A} - 1}{a - 1} > \frac{a-2}{a-1} a^{\max A}.$$

Therefore we get ( $m \in A$  implies  $m \leq \max A$ )

$$Y = \{m\} \subset A, \mu(A) < +\infty \implies \left| \int_A f d\mu \right| > \frac{a-2}{a-1} a^{\max A} \geq \frac{a-2}{a-1} a^m.$$

Since  $\lim_{m \rightarrow +\infty} \frac{a-2}{a-1} a^m = +\infty$ , we conclude that  $\int_A f d\mu = \infty$ .

If we consider the signs, then the same estimation shows that

$$\int_A f d\mu > \frac{a-2}{a-1} a^m \text{ if } m = \max A \text{ is even,}$$

and

$$\int_A f d\mu < -\frac{a-2}{a-1} a^m \text{ if } m = \max A \text{ is odd.}$$

This shows that  $\int_A f d\mu$  is neither  $+\infty$  nor  $-\infty$ .

**Exercise 10.24.** Extend Exercises 10.3 and 10.17 to the the infinity valued Lebesgue integrals.

**Exercise 10.25.** Extend the first and third properties of Propositions 10.1.4 and 10.3.6. Prove that  $f \geq g$  almost everywhere and  $\int_X g d\mu = +\infty$  implies  $\int_X f d\mu = +\infty$ .

**Exercise 10.26.** Extend the fourth property of Propositions 10.1.4 and 10.3.6.

1. Suppose  $\int_A f d\mu = +\infty$  and  $B \subset A$ . Prove that  $\int_B f d\mu$  is either finite or  $+\infty$ .
2. Suppose  $\int_A f d\mu = +\infty$  and  $\int_B f d\mu$  is either finite or  $+\infty$ . Prove that  $\int_{A \cup B} f d\mu = +\infty$ .

**Exercise 10.27.** Extend the second property of Propositions 10.1.4 and 10.3.6.

1. Prove that if  $\int_X f d\mu = +\infty$  and  $\int_X g d\mu$  is either finite or  $+\infty$ , then  $\int_X (f+g) d\mu = +\infty$ .
2. Prove that if  $\int_X f d\mu = +\infty$ , then  $\int_X c f d\mu = +\infty$  for  $c > 0$  and  $\int_X c f d\mu = -\infty$  for  $c < 0$ .

## Indefinite Integral

Suppose  $f$  is a non-negative measurable function. Then we may consider the integration on all measurable subsets

$$\nu(A) = \int_A f d\mu = \int_X f \chi_A d\mu, \quad A \in \Sigma.$$

We allow  $f$  and  $\nu$  to take extended  $+\infty$  value.

**Proposition 10.3.9.** *Suppose  $f$  is a non-negative measurable function. Then  $\nu(A) = \int_A f d\mu$  is a measure.*

*Proof.* By the fourth property of Proposition 10.3.6 and Exercise 10.26,  $\nu$  is additive. We need to show that  $\nu$  is countably additive:  $\nu(\sqcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \nu(A_i)$ . By the finite additivity of  $\nu$  and  $f \geq 0$ , it is easy to get

$$\begin{aligned} \sum_{i=1}^n \nu(A_i) &= \sum_{i=1}^n \int_{A_i} f d\mu = \int_{\sqcup_{i=1}^n A_i} f d\mu \\ &\leq \int_{\sqcup_{i=1}^n A_i} f d\mu + \int_{\sqcup_{i>n} A_i} f d\mu = \int_{\sqcup_{i=1}^{\infty} A_i} f d\mu = \nu(\sqcup_{i=1}^{\infty} A_i). \end{aligned}$$

This implies  $\sum_{i=1}^{\infty} \nu(A_i) \leq \nu(\sqcup_{i=1}^{\infty} A_i)$ . It remains to prove the inequality in the other direction.

Let  $A = \sqcup_{i=1}^{\infty} A_i$ . If  $\nu(A) = \int_A f d\mu < +\infty$ , then for any  $\epsilon > 0$ , there is  $b > 0$  and a measurable subset  $B \subset A$ , such that

$$\mu(B) < +\infty, \quad 0 \leq \int_A f d\mu - \int_B f_{[0,b]} d\mu < \epsilon.$$

Then

$$\begin{aligned} \nu(\sqcup_{i=1}^{\infty} A_i) &= \int_A f d\mu \leq \int_B f_{[0,b]} d\mu + \epsilon \\ &= \sum_{i=1}^n \int_{B \cap A_i} f_{[0,b]} d\mu + \int_{\sqcup_{i>n} (B \cap A_i)} f_{[0,b]} d\mu + \epsilon \\ &\leq \sum_{i=1}^n \int_{A_i} f d\mu + \int_{\sqcup_{i>n} (B \cap A_i)} b d\mu + \epsilon \\ &= \sum_{i=1}^n \nu(A_i) + b \mu(\sqcup_{i>n} (B \cap A_i)) + \epsilon \\ &\leq \sum_{i=1}^{\infty} \nu(A_i) + b \sum_{i>n} \mu(B \cap A_i) + \epsilon. \end{aligned}$$

By

$$\sum_{i=1}^{\infty} \mu(B \cap A_i) = \mu(B) < +\infty,$$

we see that the series  $\sum_{i=1}^{\infty} \mu(B \cap A_i)$  converges. Therefore for the known  $\epsilon$  and  $b$ , we can find big  $n$ , such that  $\sum_{i>n} \mu(B \cap A_i) < \frac{\epsilon}{b}$ . This implies

$$\nu(\sqcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \nu(A_i) + b \sum_{i>n} \mu(B \cap A_i) + \epsilon \leq \sum_{i=1}^{\infty} \nu(A_i) + 2\epsilon.$$

Since  $\epsilon$  is arbitrary, we conclude that  $\nu(\sqcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \nu(A_i)$ .

If  $\nu(A) = \int_A f d\mu = +\infty$ , then for any  $M$ , there is  $b > 0$  and a measurable subset  $B \subset A$ , such that

$$\mu(B) < +\infty, \quad \int_B f_{[0,b]} d\mu > M.$$

By the countable additivity just proved for the indefinite integral of integrable non-negative functions,  $\lambda(C) = \int_C f_{[0,b]} d\mu$  is countably additive for disjoint union of measurable subsets of  $B$ . By  $f \geq f_{[0,b]}$ , we have  $\nu(C) \geq \lambda(C)$  for  $C \subset B$ . Then by  $\sqcup_{i=1}^{\infty} (B \cap A_i) = B$ , we have

$$\sum_{i=1}^{\infty} \nu(A_i) \geq \sum_{i=1}^{\infty} \nu(B \cap A_i) \geq \sum_{i=1}^{\infty} \lambda(B \cap A_i) = \lambda(B) = \int_B f_{[0,b]} d\mu > M.$$

Since  $M$  is arbitrary, we conclude that  $\sum_{i=1}^{\infty} \nu(A_i) = +\infty$ . □

**Exercise 10.28.** Suppose  $f$  and  $g$  are extended valued measurable functions, such that  $\int_A f d\mu = \int_A g d\mu$  for any measurable subset  $A$ . Prove that if  $\mu$  is semifinite, then  $f = g$  almost everywhere. This extends Exercise 10.14 and implies the uniqueness of the “integrand” of  $\nu$  in Proposition 10.3.9 (in case  $\mu$  is semifinite).

**Exercise 10.29.** Suppose  $f$  is integrable. Prove that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\mu(A) < \delta$  implies  $\left| \int_A f d\mu \right| < \epsilon$ . Proposition 12.1.5 generalizes the exercise.

**Exercise 10.30.** Extend Exercise 10.15 to the unbounded case. Suppose  $f$  is Lebesgue integrable on any bounded interval, and the integral on the intervals always vanishes. Prove that  $f = 0$  almost everywhere.

**Exercise 10.31.** Suppose  $f$  is Lebesgue integrable on any bounded interval. Prove that  $F(x) = \int_a^x f dx$  is continuous.

## 10.4 Convergence Theorem

A key reason for introducing the Lebesgue integral is the much simpler convergence theorems. This makes the Lebesgue integral superior to the Riemann integral. The convergence theorems may be proved by using the approximation by simple functions. The idea has been used before, in the proof of Theorem 10.2.3.

### Simple Function

A function  $\phi$  is *simple* if it takes only finitely many values. Let  $c_1, \dots, c_n$  be all the distinct values of  $\phi$ , then we get a partition

$$X = \sqcup X_i, \quad X_i = \{x: \phi(x) = c_i\} = \phi^{-1}(c_i),$$

and  $\phi$  is a linear combination of characteristic functions

$$\phi = \sum_{i=1}^n c_i \chi_{X_i}.$$

Moreover,  $\phi$  is measurable if and only if  $X_i$  are measurable subsets, and we have

$$\int_X \phi d\mu = \sum c_i \mu(X_i).$$

It is also easy to see that any combination  $F(\phi_1, \dots, \phi_k)$  of finitely many simple functions is simple, and the combination is measurable if all  $\phi_i$  are measurable.

**Lemma 10.4.1.** *A non-negative function  $f$  is measurable if and only if it is the limit of an increasing sequence of non-negative measurable simple functions  $\phi_n$ .*

1. *If  $f$  is bounded, then we can further have  $\phi_n$  uniformly converging to  $f$ .*
2. *If  $f$  is integrable, then simple functions satisfying  $0 \leq \phi \leq f$  must be of the form*

$$\phi = \sum_{i=1}^n c_i \chi_{X_i}, \quad X_i \in \Sigma, \quad \mu(X_i) < +\infty, \quad c_i \neq 0, \quad (10.4.1)$$

and we can further have  $\int_X f d\mu = \lim \int_X \phi_n d\mu$ .

A consequence of the lemma is that, for a non-negative integrable function  $f$ , we have

$$\int_X f d\mu = \sup_{\phi \leq f, \phi \text{ simple}} \int_X \phi d\mu.$$

According to the Monotone Convergence Theorem (Theorem 10.4.2), the limit in the second part actually always holds as long as  $\phi_n$  is an increasing sequence of non-negative measurable simple functions converging to  $f$ .

For general  $f$ , we may apply the lemma to  $f^+$  and  $f^-$  to get similar approximations by simple functions. See Exercises 10.32 through 10.33.

If  $f$  is only equal to a measurable function almost everywhere, then we can only have  $\phi_n$  converging to  $f$  almost everywhere, and the uniform convergence only happens outside a subset of measure 0. However, this does not affect the conclusion about the Lebesgue integral.

*Proof.* If  $f$  is the limit of a sequence of measurable simple functions, then  $f$  is measurable by Proposition 10.2.2. Conversely, suppose  $f$  is measurable and non-negative. For any partition  $\Pi_n: 0 = c_0 < c_1 < \cdots < c_k = n$  of  $[0, n]$ , construct

$$\phi_n = \sum c_{i-1} \chi_{f^{-1}(c_{i-1}, c_i]}.$$

Then

$$\begin{aligned} 0 < f(x) \leq n &\implies c_{i-1} < f(x) \leq c_i \text{ for some } i \\ &\implies 0 < f(x) - \phi_n(x) \leq c_i - c_{i-1} \leq \|\Pi_n\|. \end{aligned}$$

We also note that  $f(x) = 0$  implies  $\phi_n(x) = 0$ , and  $f(x) - \phi_n(x) = 0 < \|\Pi_n\|$ .

Take a sequence of partitions  $\Pi_n$  satisfying  $\|\Pi_n\| \rightarrow 0$ . For any  $x \in X$ , we have  $0 \leq f(x) \leq N$  for some natural number  $N$ . Then

$$n > N \implies 0 \leq f(x) \leq n \implies 0 \leq f(x) - \phi_n(x) \leq \|\Pi_n\|.$$

By  $\|\Pi_n\| \rightarrow 0$ , this implies  $\lim \phi_n(x) = f(x)$ . If  $f$  is bounded, then we can find one  $N$  applicable to all  $x \in X$ , and the estimation above implies the uniform convergence. To get  $\phi_n$  to be increasing, we can either take  $\Pi_{n+1}$  to be a refinement of  $\Pi_n$ , or take the sequence  $\max\{\phi_1, \dots, \phi_n\}$  instead.

Next assume  $f$  is integrable. Suppose  $f \geq \phi = \sum_{i=1}^n c_i \chi_{X_i} \geq 0$ , and  $X_i$  are disjoint. Since  $X_i$  are disjoint, we have  $c_i \geq 0$ . By deleting those  $c_i = 0$ , we may assume that all  $c_i > 0$ . Then

$$\int_X f d\mu \geq \int_X \phi d\mu \geq \int_X c_i \chi_{X_i} d\mu = c_i \mu(X_i).$$

Since the left side is bounded and  $c_i > 0$ , we get  $\mu(X_i) < +\infty$ . Therefore  $\phi$  is of the form (12.4.1).

By the definition of  $\int_X f d\mu$ , for any  $n$ , there is a measurable  $A$  with  $\mu(A) < +\infty$  and a number  $b > 0$ , such that

$$0 \leq \int_X f d\mu - \int_A f_{[0,b]} d\mu < \frac{1}{n}.$$

By applying the first part to the bounded measurable function  $f_{[0,b]}$ , we get a simple function  $\psi_n$  satisfying  $0 \leq f_{[0,b]} - \psi_n \leq \frac{1}{n\mu(A)}$ . Then

$$0 \leq \int_X f d\mu - \int_X \psi_n d\mu \leq \left( \int_X f d\mu - \int_A f_{[0,b]} d\mu \right) + \int_A (f_{[0,b]} - \psi_n) d\mu \leq \frac{2}{n}.$$

This implies  $\int_X f d\mu = \lim \int_X \psi_n d\mu$ .

Let  $\phi_n$  be the increasing sequence of simple functions constructed before, satisfying  $0 \leq \phi_n \leq f$  and  $f = \lim \phi_n$ . Let  $\theta_n = \max\{\phi_n, \psi_1, \dots, \psi_n\}$ . Then  $\theta_n$  is an increasing sequence, and

$$0 \leq f - \theta_n \leq f - \phi_n, \quad 0 \leq f - \theta_n \leq f - \psi_n.$$

By the first inequality and  $f = \lim \phi_n$ , we get  $f = \lim \theta_n$ . By the second inequality and  $\int_X f d\mu = \lim \int_X \psi_n d\mu$ , we get  $\int_X f d\mu = \lim \int_X \theta_n d\mu$ .  $\square$

**Exercise 10.32.** Prove that a (not necessarily non-negative) function is measurable if and only if it is the limit of a sequence of measurable simple functions. Moreover, if the function is bounded, then the convergence can be uniform. Can the sequence be increasing?

**Exercise 10.33.** Suppose  $f$  is an integrable function. Prove that there is a sequence of simple functions  $\phi_n$  of the form (12.4.1), such that  $\phi_n$  converges to  $f$ , and  $\lim \int_X |f - \phi_n| d\mu = 0$ . Can  $\phi_n$  be increasing?

**Exercise 10.34.** Suppose  $\mu(X) < +\infty$  and  $f \geq 0$  is measurable. Prove that there is an increasing sequence of measurable simple functions  $\phi_n$ , such that for any  $\epsilon > 0$ , there is a measurable subset  $A$ , such that  $\mu(X - A) < \epsilon$  and  $\phi_n$  uniformly converges to  $f$  on  $A$ .

**Exercise 10.35.** In Lemma 10.4.1, prove that  $\int_X f d\mu = +\infty$  implies  $\lim \int_X \phi_n d\mu = +\infty$ .

## Convergence Theorem

The monotone limit property in Proposition 9.4.4 has the following extension.

**Theorem 10.4.2** (Monotone Convergence Theorem). *Suppose  $f_n$  is an increasing sequence of measurable functions, such that  $f_1$  is integrable. Then*

$$\int_X \lim f_n d\mu = \lim \int_X f_n d\mu.$$

The theorem (and the subsequent convergence theorems) also holds for functions that are equal to measurable functions almost everywhere.

*Proof.* By changing  $f_n$  to  $f_n - f_1$ , we may assume  $f_n \geq 0$ . Let  $f = \lim f_n$ . Then  $f$  is measurable and  $f \geq f_n$ . Therefore  $\int_X f d\mu \geq \lim \int_X f_n d\mu$ . For the converse, by Lemma 10.4.1 (and the subsequent remark), it is sufficient to prove the following: If  $\phi$  is a simple function of the form (12.4.1) satisfying  $\phi \leq f$ , then  $\int_X \phi d\mu \leq \lim \int_X f_n d\mu$ .

Fix any  $0 < \alpha < 1$ . The subset

$$X_n = \{x: f_n(x) \geq \alpha\phi(x)\}$$

is measurable and satisfies  $X_n \subset X_{n+1}$ . If  $\phi(x) = 0$ , then we always have  $f_n(x) \geq \alpha\phi(x)$ . If  $\phi(x) > 0$ , then  $\lim f_n(x) \geq \phi(x) > \alpha\phi(x)$  implies that  $x \in X_n$  for some  $n$ . Therefore  $X = \cup X_n$ . On the other hand, by applying the monotone limit property in Proposition 9.4.4 to any measurable subset  $A$  with  $\mu(A) < \infty$ , we get

$$\int_X \chi_A d\mu = \mu(A) = \lim \mu(A \cap X_n) = \lim \int_{X_n} \chi_A d\mu.$$

By taking finite linear combinations of  $\chi_A$  for such  $A$ , we get

$$\int_X \alpha\phi d\mu = \lim \int_{X_n} \alpha\phi d\mu \leq \lim \int_{X_n} f_n d\mu.$$

Since  $\alpha$  is arbitrary, we get  $\int_X \phi d\mu \leq \lim \int_{X_n} f_n d\mu$ . □

The proof also applies when  $\int_X \lim f_n d\mu = +\infty$  or  $\lim \int_X f_n d\mu = +\infty$ . Moreover, the proof shows that if  $\lim \int_X f_n d\mu$  is bounded, then  $\lim f_n$  converges almost everywhere to an integrable function.

**Theorem 10.4.3 (Fatou's Lemma).** *Suppose  $f_n$  is a sequence of measurable functions, such that  $\inf f_n$  is integrable. Then*

$$\int_X \underline{\lim} f_n d\mu \leq \underline{\lim} \int_X f_n d\mu.$$

If  $\int_X \underline{\lim} f_n d\mu = +\infty$  (this happens when  $\int_X \inf f_n d\mu = +\infty$ ), then Fatou's Lemma says  $\underline{\lim} \int_X f_n d\mu = +\infty$ .

*Proof.* Let  $g_n = \inf\{f_n, f_{n+1}, f_{n+2}, \dots\}$ . Then  $g_n$  is an increasing sequence of measurable functions, such that  $g_1$  is integrable and  $\lim g_n = \underline{\lim} f_n$ . By the Monotone Convergence Theorem, we get

$$\int_X \underline{\lim} f_n d\mu = \int_X \lim g_n d\mu = \lim \int_X g_n d\mu.$$

Moreover, by  $g_n \leq f_n$ , we get

$$\lim \int_X g_n d\mu = \underline{\lim} \int_X g_n d\mu \leq \underline{\lim} \int_X f_n d\mu. \quad \square$$

**Theorem 10.4.4** (Dominated Convergence Theorem). *Suppose  $f_n$  is a sequence of measurable functions and  $\lim f_n = f$  almost everywhere. Suppose  $|f_n| \leq g$  almost everywhere and  $g$  is integrable. Then*

$$\int_X f d\mu = \lim \int_X f_n d\mu.$$

*Proof.* The inequality  $|f_n| \leq g$  implies  $|\inf f_n| \leq g$  and  $|\sup f_n| \leq g$ . By the boundedness criterion in Proposition 10.3.5 (see Exercise 10.21), the integrability of  $g$  implies the integrability of  $\inf f_n$  and  $\sup f_n$ . Then by Fatou's Lemma,

$$\int_X \underline{\lim} f_n d\mu \leq \underline{\lim} \int_X f_n d\mu \leq \overline{\lim} \int_X f_n d\mu \leq \int_X \overline{\lim} f_n d\mu.$$

Because  $\underline{\lim} f_n = \lim f_n = \overline{\lim} f_n$ , the theorem is proved.  $\square$

**Example 10.4.1.** Consider  $f_n = -\chi_{[n,+\infty)}$ . We have  $f_n$  increasing and  $\lim f_n = 0$ . However, we also have  $\int_{\mathbb{R}} f_n dx = -\infty$  and  $\int_{\mathbb{R}} \lim f_n dx = 0$ . This shows that in the Monotone Convergence Theorem, we must have some  $f_n$  to be integrable.

**Example 10.4.2.** Example 10.4.1 also shows that Fatou's Lemma may not hold if  $\inf f_n$  is not assumed integrable. We may also consider  $f_n = -\chi_{(n,n+1)}$ , for which we have

$$\int_{\mathbb{R}} \lim f_n dx = \int_{\mathbb{R}} 0 dx = 0 \not\leq \lim \int_{\mathbb{R}} f_n dx = \lim -1 = -1.$$

In this counterexample, both sides are finite.

**Example 10.4.3.** Example 5.4.10 is a classical example that  $\lim \int f_n dx = \int \lim f_n dx$  may not hold for the Riemann integral in case  $f_n$  converges to  $f$  but not uniformly. However, the reason for the failure of the equality is only due to the non-Riemann integrability of  $\lim f_n = D$ , the Dirichlet function. In other words, the uniform convergence condition is used only for deriving the Riemann integrability of the limit.

Now the Dirichlet function is Lebesgue integrable, and  $\int D dx = 0$  because  $D = 0$  almost everywhere. Once  $\int \lim f_n dx$  makes sense, the equality is restored, by either the Monotone Convergence Theorem or the Dominated Convergence Theorem.

**Exercise 10.36.** Suppose  $f_n \geq 0$  are measurable. Prove that  $\int_X (\sum f_n) d\mu = \sum \int_X f_n d\mu$ .

**Exercise 10.37.** Show that it is possible to have strict inequality in Fatou's Lemma.

**Exercise 10.38.** Suppose  $f_n$  are measurable,  $f_1 + \cdots + f_n \geq 0$ ,  $\sum f_n$  converges, and  $\sum \int_X f_n d\mu$  converges. Prove that  $\int_X (\sum f_n) d\mu \leq \sum \int_X f_n d\mu$ . Moreover, show that the strict inequality may happen.



Exercise 10.39. Prove Proposition 10.3.9 by using Monotone Convergence Theorem.

Exercise 10.40. Redo Exercise 10.31 by using Dominated Convergence Theorem.

Exercise 10.41. Find a counterexample for the Dominated Convergence Theorem.

Exercise 10.42 (Dominated Convergence Theorem, extended). Suppose  $f_n$  and  $f$  are measurable. Suppose  $g_n$  and  $g$  are integrable. If  $\lim f_n = f$ ,  $\lim g_n = g$ ,  $|f_n| \leq g_n$  and  $\lim \int_X g_n d\mu = \int_X g d\mu$ , then

$$\lim \int_X f_n d\mu = \int_X f d\mu.$$

Exercise 10.43. Suppose  $f_n$  and  $f$  are integrable and  $\lim f_n = f$ . Prove that  $\lim \int_X |f_n - f| d\mu = 0$  if and only if  $\lim \int_X |f_n| d\mu = \int_X |f| d\mu$ .

Exercise 10.44. Suppose  $f(x, t)$  is measurable for each fixed  $t$ . Suppose  $\lim_{t \rightarrow 0} f(x, t) = f(x)$  and  $|f(x, t)| \leq g(x)$  for some integrable  $g$ . Prove that

$$\lim_{t \rightarrow 0} \int f(x, t) dx = \int f(x) dx.$$

Moreover, prove that if  $f$  is continuous in  $t$ , then  $\int f(x, t) dx$  is continuous in  $t$ .

Exercise 10.45. Suppose  $f(x, t)$  is measurable for each fixed  $t$ . Suppose the partial derivative  $\frac{\partial f}{\partial t}$  exists and is bounded. Prove that for any bounded  $[a, b]$ ,

$$\frac{d}{dt} \int_a^b f(x, t) dx = \int_a^b \frac{\partial f}{\partial t}(x, t) dx.$$

Exercise 10.46. Prove Borel-Cantelli Lemma in Exercise 9.50 by applying the Monotone Convergence Theorem to the integration of the series  $\sum \chi_{A_n}$ .

## Criterion for Riemann Integrability

A consequence of the Monotone Convergence Theorem is the following criterion for Riemann integrability.

**Theorem 10.4.5 (Lebesgue Theorem).** *A bounded function on  $[a, b]$  is Riemann integrable if and only if it is continuous almost everywhere.*

*Proof.* Suppose  $f$  is Riemann integrable. In Exercise 4.85, we introduced the oscillation of  $f$  at  $x \in [a, b]$

$$\omega_x(f) = \lim_{\epsilon \rightarrow 0^+} \omega_{(x-\epsilon, x+\epsilon)} f,$$

and showed that  $f$  is continuous at  $x$  if and only if  $\omega_x(f) = 0$ . So  $A = \{x: \omega_x(f) > 0\}$  is exactly the collection of places where  $f$  is not continuous. We have

$$A = \cup A_{\frac{1}{n}}, \quad A_\delta = \{x: \omega_x(f) > \delta\}.$$

Thus it is sufficient to prove that  $A_\delta$  is a subset of measure 0 for any  $\delta > 0$ .

For any  $\epsilon > 0$ , there is a partition  $P$  of  $[a, b]$  by intervals, such that

$$\sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i < \epsilon \delta.$$

If  $x \in (x_{i-1}, x_i)$  and  $\omega_x(f) > \delta$ , then  $\omega_{[x_{i-1}, x_i]}(f) \geq \omega_x(f) > \delta$ . Therefore

$$A_\delta \subset \sqcup_{\omega_{[x_{i-1}, x_i]}(f) > \delta} (x_{i-1}, x_i) \sqcup \{x_0, x_1, \dots, x_n\}.$$

On the other hand,

$$\begin{aligned} \epsilon \delta &> \sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i \geq \sum_{\omega_{[x_{i-1}, x_i]}(f) > \delta} \omega_{[x_{i-1}, x_i]}(f) \Delta x_i \\ &\geq \delta \mu(\sqcup_{\omega_{[x_{i-1}, x_i]}(f) > \delta} (x_{i-1}, x_i)). \end{aligned}$$

Therefore

$$\mu(\sqcup_{\omega_{[x_{i-1}, x_i]}(f) > \delta} (x_{i-1}, x_i) \sqcup \{x_0, x_1, \dots, x_n\}) = \mu(\sqcup_{\omega_{[x_{i-1}, x_i]}(f) > \delta} (x_{i-1}, x_i)) < \epsilon.$$

Since  $\epsilon$  is arbitrary,  $A_\delta$  is a subset of measure zero.

Conversely, suppose  $f$  is continuous almost everywhere. Let  $P_n$  be a sequence of partitions of  $[a, b]$  by intervals, such that  $P_{n+1}$  refines  $P_n$  and  $\|P_n\| \rightarrow 0$ . For each  $P_n$ , let

$$\phi_n = \sum \left( \inf_{[x_{i-1}, x_i]} f \right) \chi_{(x_{i-1}, x_i]}, \quad \psi_n = \sum \left( \sup_{[x_{i-1}, x_i]} f \right) \chi_{(x_{i-1}, x_i]}.$$

Since  $P_{n+1}$  refines  $P_n$ ,  $\phi_n$  is decreasing and  $\psi_n$  is increasing. Since  $\|P_n\| \rightarrow 0$ , we have  $\lim \phi_n(x) = \lim \psi_n(x)$  whenever  $f$  is continuous at  $x$ . Therefore we get  $\lim \phi_n = \lim \psi_n$  almost everywhere. By the Monotone Convergence Theorem, we have

$$\lim \int_a^b \phi_n d\mu = \int_a^b \lim \phi_n d\mu = \int_a^b \lim \psi_n d\mu = \lim \int_a^b \psi_n d\mu.$$

On the other hand, we know

$$\int_a^b \psi_n d\mu - \int_a^b \phi_n d\mu = \sum \left( \sup_{[x_{i-1}, x_i]} f - \inf_{[x_{i-1}, x_i]} f \right) \Delta x_i = \sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i.$$

Thus the criterion for the Riemann integrability is verified.  $\square$

**Exercise 10.47.** Suppose  $f$  is a bounded function on  $[a, b]$ . Prove that the integral of the lower (upper) envelope of  $f$  in Exercise 10.12 is the lower (upper) Darboux integral of  $f$

$$\int_a^b f_*(x) dx = \int_a^b f(x) dx, \quad \int_a^b f^*(x) dx = \int_a^b f(x) dx.$$

Then use this to prove Lebesgue theorem.

## 10.5 Convergence and Approximation

We have used the approximation by simple functions. The approximation by other good functions can also be useful. On the other hand, we saw two meanings of the approximation in Lemma 10.4.1. A sequence  $f_n$  uniformly converges to  $f$  if

$$\lim \|f_n - f\|_\infty = 0,$$

where  $\|f\|_\infty = \sup_{x \in X} |f(x)|$  is the  $L^\infty$ -norm. We also have the convergence

$$\lim \|f_n - f\|_1 = 0,$$

with respect to the  $L^1$ -norm  $\|f\|_1 = \int_X |f| d\mu$ . This implies the convergence

$\lim \int_X f_n d\mu = \int_X f d\mu$  of the Lebesgue integral. Another useful notion is *convergence in measure*, which means that for any  $\epsilon > 0$ , we have

$$\lim_{n \rightarrow \infty} \mu(\{x: |f_n(x) - f(x)| \geq \epsilon\}) = 0.$$

### Almost Uniform Convergence

**Proposition 10.5.1** (Egorov's Theorem). *Suppose  $f_n$  is a sequence of measurable function converging to  $f$  almost everywhere. If  $\mu(X) < +\infty$ , then for any  $\epsilon > 0$ ,  $f_n$  converges uniformly to  $f$  outside a subset of measure  $< \epsilon$ .*

*Proof.* Let  $g_n = f_n - f$  and let  $\epsilon_n > 0$  converge to 0. Then  $g_n$  converges to 0 almost everywhere. This means that for almost all  $x \in X$  and any  $\epsilon_m$ , there is  $N$ , such that  $n > N$  implies  $|g_n(x)| < \epsilon_m$ . In other words, almost all  $x$  belongs to the subset  $Y = \bigcap_{m=1}^\infty \bigcup_{N=1}^\infty \bigcap_{n>N} \{x: |g_n(x)| < \epsilon_m\}$ . Let

$$E_{m,N} = \bigcup_{n>N} \{x: |g_n(x)| \geq \epsilon_m\} = \{x: |g_n(x)| \geq \epsilon_m \text{ for some } n > N\}.$$

Then  $\bigcup_{m=1}^\infty \bigcap_{N=1}^\infty E_{m,N}$  is the complement of  $Y$ , so that  $\mu(\bigcup_{m=1}^\infty \bigcap_{N=1}^\infty E_{m,N}) = 0$ . This is the same as  $\mu(\bigcap_{N=1}^\infty E_{m,N}) = 0$  for any  $m$ . Since  $\mu(X) < +\infty$  and  $E_{m,N}$  is decreasing in  $N$ , by the monotone limit property in Proposition 9.4.4, for fixed  $m$ , we have  $\lim_{N \rightarrow \infty} \mu(E_{m,N}) = 0$ .

On the other hand, the uniform convergence of  $f_n$  on a subset  $A$  means that for any  $x \in A$  and any  $\epsilon_m$ , there is  $N$ , such that  $n > N$  implies  $|g_n(x)| < \epsilon_m$ . In other words, for each  $\epsilon_m$ , there is  $N$ , such that  $A \subset \bigcap_{n>N} \{x: |g_n(x)| < \epsilon_m\} = X - E_{m,N}$ . By this interpretation, for any sequence  $N_m$ , the sequence  $f_n$  uniformly converges on  $\bigcap_{m=1}^\infty (X - E_{m,N_m}) = X - \bigcup_{m=1}^\infty E_{m,N_m}$ .

It remains to find a sequence  $N_m$ , such that  $\mu(\bigcup_{m=1}^\infty E_{m,N_m})$  is as small as we wish. For any given  $\epsilon > 0$  and  $m$ , by  $\lim_{N \rightarrow \infty} \mu(E_{m,N}) = 0$ , we can find  $N_m$ , such that  $\mu(E_{m,N_m}) < \frac{\epsilon}{2^m}$ . Then  $\mu(\bigcup_{m=1}^\infty E_{m,N_m}) \leq \sum_{m=1}^\infty \mu(E_{m,N_m}) < \sum_{m=1}^\infty \frac{\epsilon}{2^m} = \epsilon$ .  $\square$

**Exercise 10.48.** Show that Egorov's Theorem fails on the whole  $\mathbb{R}$  with the usual Lebesgue measure. Therefore the condition  $\mu(X) < +\infty$  is necessary.

**Exercise 10.49.** Use Lemma 10.4.1 and Egorov's Theorem to give another proof of Exercise 10.34.

**Exercise 10.50.** Prove the converse of Egorov's Theorem: Suppose for any  $\epsilon > 0$ , there is a subset  $A$ , such that  $\mu(X - A) < \epsilon$  and  $f_n$  converges (not necessarily uniformly) to  $f$  on  $A$ , then  $f_n$  converges to  $f$  almost everywhere.

## Convergence in Measure

**Proposition 10.5.2.** *Suppose  $f_n$  is a sequence of measurable functions. If  $f_n$  converges almost everywhere and  $\mu(X) < +\infty$ , then  $f_n$  converges in measure. Conversely, if  $f_n$  converges in measure, then a subsequence converges almost everywhere.*

*Proof.* Let  $f_n$  converge to  $f$  almost everywhere and let  $\epsilon > 0$  be given. By  $\mu(X) < +\infty$ , we may apply Egorov's Theorem. For any  $\delta > 0$ , there is a measurable subset  $A$ , such that  $\mu(A) < \delta$  and  $f_n$  uniformly converges to  $f$  on  $X - A$ . The uniform convergence means that, for the given  $\epsilon > 0$ , there is  $N$ , such that

$$x \in X - A, n > N \implies |f_n(x) - f(x)| < \epsilon.$$

This means that

$$\begin{aligned} n > N &\implies \{x: |f_n(x) - f(x)| \geq \epsilon\} \subset A \\ &\implies \mu(\{x: |f_n(x) - f(x)| \geq \epsilon\}) \leq \mu(A) < \delta. \end{aligned}$$

This proves  $\lim_{n \rightarrow \infty} \mu(\{x: |f_n(x) - f(x)| \geq \epsilon\}) = 0$ .

Conversely, let  $f_n$  converge to  $f$  in measure and let  $\epsilon_k > 0$  be a sequence, such that  $\sum \epsilon_k$  converges. By  $\lim \mu(\{x: |f_n(x) - f(x)| \geq \epsilon_k\}) = 0$ , for any  $\epsilon_k$ , there is  $n_k$ , such that  $\mu(\{x: |f_{n_k}(x) - f(x)| \geq \epsilon_k\}) \leq \epsilon_k$ .

Suppose  $x \notin A_K = \cup_{k > K} \{x: |f_{n_k}(x) - f(x)| \geq \epsilon_k\}$  for some  $K$ , then  $k > K$  implies  $|f_{n_k}(x) - f(x)| < \epsilon_k$ . By  $\epsilon_k \rightarrow 0$ , this further implies  $\lim f_{n_k}(x) = f(x)$ . Therefore we have  $\lim f_{n_k}(x) = f(x)$  for  $x \notin \cap_{K=1}^{\infty} A_K$ . It remains to show that  $\mu(\cap_{K=1}^{\infty} A_K) = 0$ .

The sequence  $A_K$  is decreasing and satisfies

$$\mu(A_K) \leq \sum_{k > K} \mu(\{x: |f_{n_k}(x) - f(x)| \geq \epsilon_k\}) \leq \sum_{k > K} \epsilon_k.$$

Since the right side converges to 0 as  $K \rightarrow \infty$ , by the monotone limit property in Proposition 9.4.4, we have  $\mu(\cap_{K=1}^{\infty} A_K) = \lim_{K \rightarrow \infty} \mu(A_K) = 0$ .  $\square$

**Example 10.5.1.** Convergence in measure does not necessarily imply convergence almost everywhere.

Consider the sequence of subsets  $I_n \subset [0, 1]$  given by

$$[0, 1], \left[0, \frac{1}{2}\right], \left[\frac{1}{2}, 1\right], \left[0, \frac{1}{3}\right], \left[\frac{1}{3}, \frac{2}{3}\right], \left[\frac{2}{3}, 1\right], \left[0, \frac{1}{4}\right], \left[\frac{1}{4}, \frac{2}{4}\right], \left[\frac{2}{4}, \frac{3}{4}\right], \left[\frac{3}{4}, 1\right], \dots$$

Let  $f_n = \chi_{I_n}$ . For any  $0 < \epsilon < 1$ , we have  $\{x: |f_n(x)| \geq \epsilon\} \subset I_n$ . Therefore  $f_n$  converges to 0 in measure. On the other hand, for any  $x \in [0, 1]$ , there are subsequences  $f_{m_k}$  and  $f_{n_k}$ , such that  $f_{m_k}(x) = 0$  and  $f_{n_k}(x) = 1$ . Therefore the sequence diverges everywhere.

**Exercise 10.51.** Prove that if  $\lim_{n \rightarrow \infty} \int_X |f_n - f| d\mu = 0$ , then  $f_n$  converges to  $f$  in measure. Is the converse true?

### Approximation by Smooth Function

A function on  $\mathbb{R}$  is *smooth* if it has derivatives of any order. It is *compactly supported* if it is zero outside a big interval. The notions can be extended to multivariable functions.

**Proposition 10.5.3.** *Suppose  $f$  is Lebesgue integrable on  $\mathbb{R}$ . Then for any  $\epsilon > 0$ , there is a compactly supported smooth function  $g$ , such that  $\int |f - g| dx < \epsilon$ .*

*Proof.* For the special case  $f = \chi_{(a,b)}$  is the characteristic function of a bounded interval, Figure 4.3.1 shows that, for any  $\epsilon > 0$ , it is not difficult to construct a compactly supported smooth function  $g$ , such that  $\int |\chi_{(a,b)} - g| dx < \epsilon$ .

Now consider the case  $f = \chi_A$  is the characteristic function of a Lebesgue measurable subset  $A$  of finite measure. For any  $\epsilon > 0$ , we have  $\mu(U - A) < \epsilon$  for some open  $U \supset A$ . Let  $U = \sqcup_{i=1}^{\infty} (a_i, b_i)$ . Then  $\sum (b_i - a_i)$  converges, so that there is  $n$ , such that  $V = \sqcup_{i=1}^n (a_i, b_i)$  satisfies  $\mu(U - V) < \epsilon$ . Moreover,

$$\mu((V - A) \cup (A - V)) \leq \mu(U - A) + \mu(U - V) < 2\epsilon.$$

We have  $\chi_V = \sum \chi_{(a_i, b_i)}$ . There are compactly supported smooth function  $g_i$  satisfying  $\int |\chi_{(a_i, b_i)} - g_i| dx < \frac{\epsilon}{n}$ . Then  $g = \sum_{i=1}^n g_i$  is a compactly supported smooth function satisfying

$$\begin{aligned} \int |\chi_A - g| dx &\leq \int |\chi_A - \chi_V| dx + \int |\chi_V - g| dx \\ &\leq \mu((V - A) \cup (A - V)) + \sum_{i=1}^n \int |\chi_{(a_i, b_i)} - g_i| dx \\ &< 2\epsilon + \sum_{i=1}^n \frac{\epsilon}{n} = 3\epsilon. \end{aligned}$$

Next consider integrable  $f$ . By Exercise 10.33, for any  $\epsilon > 0$ , there is a simple function  $\phi = \sum_{i=1}^n c_i \chi_{A_i}$ , such that  $\mu(A_i) < +\infty$  and  $\int |f - \phi| dx < \epsilon$ . We have shown that there are compactly supported smooth functions  $g_i$  satisfying  $\int |\chi_{A_i} - g_i| dx < \frac{\epsilon}{n|c_i|}$ . Then  $g = \sum c_i g_i$  is a compactly supported smooth function satisfying

$$\int |f - g| dx \leq \int |f - \phi| dx + \sum |c_i| \int |\chi_{A_i} - g_i| dx < 2\epsilon. \quad \square$$

The proof above is quite typical, in that the proof is carried out first for the characteristic function of an interval, then for the characteristic function of a measurable subset. By taking linear combinations, we get the proof for simple functions. By approximation by simple functions, the proof is further extended to Lebesgue integrable functions.

**Exercise 10.52 (Riemann-Lebesgue Lemma).** Suppose  $f(x)$  is a bounded Lebesgue measurable periodic function with period  $T$ . Suppose  $g(x)$  is Lebesgue integrable on  $\mathbb{R}$ . Prove that

$$\lim_{t \rightarrow \infty} \int_a^b f(tx)g(x)dx = \frac{1}{T} \int_0^T f(x)dx \int_a^b g(x)dx.$$

The proof may start for the case  $g$  is the characteristic function of an interval. Also compare with Exercise 4.38.

**Exercise 10.53.** Suppose  $f(x)$  is a bounded Lebesgue measurable function and  $g(x)$  is Lebesgue integrable on  $\mathbb{R}$ . Prove that  $\lim_{t \rightarrow 0} \int f(x)|g(x) - g(x+t)|dx = 0$ .

## Testing Function

Two integrable functions  $f_1$  and  $f_2$  are equal almost everywhere if and only if  $\int_A f_1 d\mu = \int_A f_2 d\mu$  for any measurable  $A$ . This is the same as  $\int_X (f_1 - f_2)\chi_A d\mu = 0$ . By approximating  $\chi_A$  with nice classes of functions  $g$ , the criterion becomes  $\int_X f_1 g d\mu = \int_X f_2 g d\mu$  for all nice  $g$ .

**Proposition 10.5.4.** Suppose  $f$  is Lebesgue integrable on  $\mathbb{R}$ . If  $\int f g d\mu = 0$  for any compactly supported smooth function  $g$ , then  $f = 0$  almost everywhere.

The earlier Example 4.3.6 also illustrates the idea of testing function.

*Proof.* For any  $\epsilon > 0$ , there is  $b$ , such that  $\int |f - f_{[-b,b]}|d\mu < \epsilon$ . By Proposition 10.5.3, for any bounded Lebesgue measurable subset  $A$  and  $\epsilon > 0$ , there is a compactly supported smooth function  $g$ , such that  $\int |\chi_A - g|d\mu < \frac{\epsilon}{b}$ . In fact, by tracing the construction of  $g$ , we may further assume  $|g| \leq 1$ . Then

$$\begin{aligned} \left| \int_A f d\mu \right| &\leq \left| \int f_{[-b,b]}(\chi_A - g) d\mu \right| + \int_A |f - f_{[-b,b]}| d\mu \\ &< b \int |\chi_A - g| d\mu + \int |f - f_{[-b,b]}| d\mu < 2\epsilon. \end{aligned}$$

Since  $\epsilon$  can be arbitrarily small, we get  $\int_A f d\mu = 0$  for any bounded Lebesgue measurable  $A$ . This implies  $f = 0$  almost everywhere.  $\square$

### Approximation by Continuous Function

**Proposition 10.5.5** (Lusin's Theorem). *Suppose  $f$  is a Lebesgue measurable function on  $\mathbb{R}$ . Then for any  $\epsilon > 0$ , there is a continuous function  $g$  on  $\mathbb{R}$ , such that  $f(x) = g(x)$  for  $x$  outside a subset of measure  $< \epsilon$ .*

*Proof.* For the case  $f = \chi_A$  is the characteristic function of a Lebesgue measurable subset  $A$ , we have closed  $K$  and open  $U$ , such that  $K \subset A \subset U$  and  $\mu(U - K) < \epsilon$ . For the closed subsets  $K$  and  $\mathbb{R} - U$ , the distance functions

$$d(x, K) = \inf\{|x - c| : c \in K\}, \quad d(x, \mathbb{R} - U) = \inf\{|x - c| : c \in \mathbb{R} - U\}$$

are continuous and satisfy

$$d(x, K) = 0 \iff x \in K, \quad d(x, \mathbb{R} - U) = 0 \iff x \in \mathbb{R} - U.$$

Then the fact that  $K$  and  $\mathbb{R} - U$  are disjoint implies  $d(x, K) + d(x, \mathbb{R} - U) > 0$  for all  $x$ , and therefore the function

$$g(x) = \frac{d(x, \mathbb{R} - U)}{d(x, K) + d(x, \mathbb{R} - U)}$$

is defined and continuous on the whole  $\mathbb{R}$ . Moreover, we have  $\chi_A = 1 = g$  on  $K$  and  $\chi_A = 0 = g$  on  $\mathbb{R} - U$ . The place the two functions are different lies in  $\mathbb{R} - K - (\mathbb{R} - U) = U - K$ , and we have  $\mu(U - K) < \epsilon$ .

For a measurable simple function  $\phi = \sum_{i=1}^n c_i \chi_{A_i}$ , we can find continuous  $g_i$ , such that  $\chi_{A_i} = g_i$  outside a subset of measure  $< \frac{\epsilon}{n}$ . Then  $g = \sum c_i g_i$  is a continuous function, such that  $\phi = g$  outside a subset of measure  $< \epsilon$ . We also note that, if  $a \leq \phi \leq b$ , then the truncation  $g_{[a,b]}$  is also a continuous function satisfying  $\phi = g_{[a,b]}$  whenever  $\phi = g$ . Therefore we may also assume  $a \leq g \leq b$ .

Now suppose  $f$  is a bounded Lebesgue measurable function. By Lemma 10.4.1, there is an increasing sequence of simple functions  $\phi_n$  uniformly converging to  $f$ . In fact, the proof of the lemma shows that, if  $\delta_n > 0$  are fixed, such that  $\sum \delta_n$  converges, then it is possible to arrange to have  $0 \leq \phi_n - \phi_{n-1} \leq \delta_n$ . We have shown that each simple function  $\phi_n - \phi_{n-1}$  is equal to a continuous function  $g_n$  outside a subset  $A_n$  of measure  $< \frac{\epsilon}{2^n}$ . Moreover, we may arrange to have  $0 \leq g_n \leq \delta_n$ . The convergence of  $\sum \delta_n$  implies the uniform convergence of  $\sum g_n$ , so that  $g = \sum g_n$  is continuous. Moreover, we have  $f = \sum(\phi_n - \phi_{n-1}) = \sum g_n = g$  outside the subset  $\cup A_n$ , which has measure  $\mu(\cup A_n) \leq \sum \mu(A_n) < \sum \frac{\epsilon}{2^n} = \epsilon$ .

Next for any Lebesgue measurable function  $f$ , we consider the restriction  $f\chi_{(a,b)}$  of the function to a bounded interval  $(a, b)$ . By Exercise 10.8, for any  $\epsilon > 0$ , there is  $A \subset (a, b)$ , such that  $\mu((a, b) - A) < \epsilon$  and  $f$  is bounded on  $A$ . Then  $f\chi_A$  is a bounded Lebesgue measurable function. By the earlier part of the proof, we have  $f\chi_A = g$  on  $B$  for a continuous function  $g$  and a subset  $B \subset (a, b)$  satisfying  $\mu((a, b) - B) < \epsilon$ . Then  $f = f\chi_A = g$  on  $A \cap B$ , with  $\mu((a, b) - A \cap B) \leq \mu((a, b) - A) + \mu((a, b) - B) < 2\epsilon$ .

It remains to combine the restrictions of  $f$  to bounded intervals together. We cover  $\mathbb{R}$  by countably many bounded intervals:  $\mathbb{R} = \cup_{i=1}^{\infty} (a_i, b_i)$ , such that any  $x \in \mathbb{R}$  has a neighborhood  $(x - \delta, x + \delta)$  intersecting only finitely many  $(a_i, b_i)$  (this is called *locally finite* property). We also find continuous functions  $\alpha_i$  satisfying

$$\alpha_i \geq 0, \quad \sum \alpha_i = 1, \quad \alpha_i = 0 \text{ on } \mathbb{R} - (a_i, b_i).$$

The functions may be constructed by taking  $\beta_i$  as in Figure 4.3.1 ( $c - 2\delta$  and  $c + 2\delta$  are respectively  $a_i$  and  $b_i$ ), which satisfies

$$\beta_i > 0 \text{ on } (a_i, b_i), \quad \beta_i = 0 \text{ on } \mathbb{R} - (a_i, b_i),$$

and then taking

$$\alpha_i = \frac{\beta_i}{\sum \beta_i}.$$

We note that by the locally finite property, the sum  $\sum \beta_i$  is always a finite sum on a neighborhood of any  $x \in \mathbb{R}$  and is therefore continuous. Moreover, the covering  $\mathbb{R} = \cup_{i=1}^{\infty} (a_i, b_i)$  implies that  $\sum \beta_i > 0$  everywhere. By what we just proved before, for each  $(a_i, b_i)$ , there is a continuous function  $g_i = f$  on  $A_i \subset (a_i, b_i)$ , with  $\mu((a_i, b_i) - A_i) < 2^{-i}\epsilon$ . Just like the continuity of  $\sum \beta_i$ , the sum  $g = \sum \alpha_i g_i$  is also continuous. By comparing  $g = \sum \alpha_i g_i$  and  $f = f \sum \alpha_i = \sum \alpha_i f$ , we find that  $g(x) \neq f(x)$  implies  $\alpha_i(x)g_i(x) \neq \alpha_i(x)f(x)$  for some  $i$ . This further implies  $x \in (a_i, b_i)$  and  $g_i(x) \neq f(x)$ , and therefore  $x \notin A_i$ . We conclude that the places where  $g \neq f$  is contained in  $\cup((a_i, b_i) - A_i)$ , and  $\mu(\cup((a_i, b_i) - A_i)) \leq \sum \mu((a_i, b_i) - A_i) < \epsilon$ .  $\square$

The collection of functions  $\alpha_i$  is a *partition of unity* with respect to the (locally finite) covering  $\mathbb{R} = \cup_{i=1}^{\infty} (a_i, b_i)$ . The proof above is a typical example of using the partition of unity to combine the constructions on the pieces in a covering to form a global construction.

Note that it is possible to find smooth  $\beta_i$ . Then the locally finite property implies that the functions  $\alpha_i$  are also smooth.

**Exercise 10.54.** Use Propositions 10.5.1, 10.5.2, 10.5.3, and Exercise 10.51 to give another proof of Lusin's Theorem.

## 10.6 Additional Exercise

### Darboux Theory of the Lebesgue Integral

Suppose  $f$  is a bounded function on a measure space  $(X, \Sigma, \mu)$  with finite  $\mu(X)$ . Define

$$\int_X f d\mu = \sup_{\phi \leq f} \sum c_i \mu(X_i), \quad \overline{\int}_X f d\mu = \inf_{\phi \geq f} \sum c_i \mu(X_i)$$

where  $\phi = \sum_{i=1}^n c_i \chi_{X_i}$  are simple functions.

**Exercise 10.55.** Prove that  $\int_X f d\mu = \sup_P L(P, f)$ , where  $L(P, f)$  is introduced in the proof of Proposition 10.1.2.



Exercise 10.56. Prove that  $f$  is integrable if and only if  $\int_{\underline{X}} f d\mu = \overline{\int_X f d\mu}$ .

### Lebesgue Integral of Vector Valued Function: Bounded Case

Let  $(X, \Sigma, \mu)$  be a measure space with  $\mu(X) < +\infty$ . Let  $F: X \rightarrow \mathbb{R}^n$  be a bounded map.

Exercise 10.57. Define the Lebesgue integral  $\int_X F d\mu$  in terms of the “Riemann sum”.

Exercise 10.58. Extend the integrability criterion in Proposition 10.1.2 to  $\int_X F d\mu$ .

Exercise 10.59. Prove that  $F$  is Lebesgue integrable if and only if all its coordinate functions are Lebesgue integrable. Moreover, the coordinates of  $\int_X F d\mu$  are the Lebesgue integrals of the coordinate functions.

Exercise 10.60. Extend the properties of the Lebesgue integration in Proposition 10.1.4 to vector valued functions.

Exercise 10.61. Prove that the Lebesgue integrability of  $F$  implies the Lebesgue integrability of  $\|F\|$ , and  $\left\| \int_X F d\mu \right\| \leq \int_X \|F\| d\mu$ . This is the partial extension of the fourth property in Proposition 10.1.4.

Exercise 10.62. Let  $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear transform. Prove that if  $F$  is Lebesgue integrable, then  $L \circ F$  is also Lebesgue integrable, and  $\int_X L \circ F d\mu = L \circ \int_X F d\mu$ .

### Measurable Vector Valued Function

Let  $\Sigma$  be a  $\sigma$ -algebra on a set  $X$ . A map  $F: X \rightarrow \mathbb{R}^n$  is *measurable* if  $F^{-1}(I) \in \Sigma$  for any rectangle  $I = (a_1, b_1] \times \cdots \times (a_n, b_n]$ .

Exercise 10.63. Show that the measurability can be defined in terms of other types of rectangles. The rectangles can also be replaced by open, closed, compact, or Borel subsets of  $\mathbb{R}^n$ . See Definition 11.4.2 and the subsequent discussion.

Exercise 10.64. Prove that  $F$  is measurable if and only if all its coordinate functions are measurable.

Exercise 10.65. For maps  $F: X \rightarrow \mathbb{R}^n$  and  $G: X \rightarrow \mathbb{R}^m$ , prove that  $(F, G): X \rightarrow \mathbb{R}^{m+n}$  is measurable if and only if  $F$  and  $G$  are measurable.

Exercise 10.66. If  $X = \cup X_i$  is a countable union and  $X_i \in \Sigma$ , prove that  $F$  is measurable if and only if the restrictions  $F|_{X_i}$  are measurable. The extension of the other properties in Proposition 10.2.2 are also true (we can only talk about  $\lim F_n$  in the third property), by using open subsets in Exercise 10.63.

**Exercise 10.67.** Suppose  $F$  is bounded and  $\mu$  is a measure on  $(X, \Sigma)$  satisfying  $\mu(X) < +\infty$ . Prove that  $F$  is Lebesgue integrable if and only if it is equal to a measurable map almost everywhere.

### Lebesgue Integral of Vector Valued Function: General Case

Let  $(X, \Sigma, \mu)$  be a measure space. Let  $F: X \rightarrow \mathbb{R}^n$  be a measurable map.

**Exercise 10.68.** Define the Lebesgue integral  $\int_X F d\mu$ .

**Exercise 10.69.** Prove that  $F$  is Lebesgue integrable if and only if all its coordinate functions are Lebesgue integrable. Moreover, the coordinates of  $\int_X F d\mu$  are the Lebesgue integrals of the coordinate functions.

**Exercise 10.70.** Extend Exercises 10.60, 10.61, 10.62 to general vector valued functions.

**Exercise 10.71.** Prove that  $\lim_{\sqcup X_i} \sum_i \left\| \int_{X_i} F d\mu \right\| = \int_X \|F\| d\mu$ , where the limit is with respect to the refinements of measurable partitions.

**Exercise 10.72.** Formulate and prove the comparison test for the integrability of vector valued functions.

**Exercise 10.73.** Formulate and prove the Dominated Convergence Theorem for vector valued functions.

## Chapter 11

# Product Measure

## 11.1 Extension Theorem

The natural generalization of intervals to Euclidean spaces is the rectangles

$$I = \langle a_1, b_1 \rangle \times \langle a_2, b_2 \rangle \times \cdots \times \langle a_n, b_n \rangle.$$

The high dimensional version of the Lebesgue measure should extend the volume

$$\lambda(I) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n).$$

However, the old argument for the Lebesgue measure on  $\mathbb{R}$  cannot be extended because open subsets in Euclidean spaces are not necessarily disjoint unions of open rectangles. Instead of using open subsets to define the outer measure, we need to define the outer measure directly from the volume of rectangles.

### Outer Measure Induced by Special Volume

**Proposition 11.1.1.** *Suppose  $\mathcal{C}$  is a collection of subsets in  $X$  and  $\lambda$  is a non-negative extended valued function on  $\mathcal{C}$ . Define*

$$\mu^*(A) = \inf \left\{ \sum \lambda(C_i) : A \subset \cup C_i, C_i \in \mathcal{C} \right\}$$

*in case  $A$  is contained in a countable union of subsets in  $\mathcal{C}$ , and define  $\mu^*(A) = +\infty$  otherwise. Then  $\mu^*$  is an outer measure.*

We consider an empty union to be the empty set, and take the corresponding “empty sum”  $\sum \lambda(C_i)$  to be 0. Therefore  $\lambda(\emptyset) = 0$  is implicit in the definition.

*Proof.* The definition clearly satisfies the monotone property. The countable subadditivity is also obviously satisfied when some  $A_i$  is not contained in some countable union of subsets in  $\mathcal{C}$ .

So we assume each  $A_i$  is contained in some countable union of subsets in  $\mathcal{C}$  and try to prove the subadditivity. The proof is essentially the same as the proof of the subadditivity property in Proposition 9.2.2.

For any  $\epsilon > 0$  and  $i \in \mathbb{N}$ , there is  $\cup_j C_{ij} \supset A_i$ ,  $C_{ij} \in \mathcal{C}$ , such that

$$\sum_j \lambda(C_{ij}) \leq \mu^*(A_i) + \frac{\epsilon}{2^i}.$$

By  $\cup A_i \subset \cup_{ij} C_{ij}$ , we get

$$\mu^*(\cup A_i) \leq \sum_{ij} \lambda(C_{ij}) = \sum_i \left( \sum_j \lambda(C_{ij}) \right) \leq \sum_i \mu^*(A_i) + \epsilon.$$

Because  $\epsilon$  is arbitrary, we conclude the countable subadditivity.  $\square$

**Example 11.1.1.** Let  $\mathcal{C}$  be the collection of single point subsets in  $X$  and  $\lambda\{x\} = 1$  for any  $x \in X$ . Then the outer measure in Proposition 11.1.1 is the counting (outer) measure in Examples 9.3.2 and 9.3.5.

**Example 11.1.2.** We fix  $a \in X$  and define  $\lambda\{x\} = 1$  if  $x = a$  and  $\lambda\{x\} = 0$  otherwise. Then  $\lambda$  induces the outer measure that in turn gives the Dirac measure in Example 9.4.5.

**Example 11.1.3.** Let  $\mathcal{C}$  be the collection of two element subsets of  $X$  and  $\lambda\{x, y\} = 1$  for any  $\{x, y\} \in \mathcal{C}$ . Then  $\lambda$  induces the outer measure

$$\mu^*(A) = \begin{cases} \frac{|A|}{2}, & \text{if } A \text{ contains even number of elements,} \\ \frac{|A|+1}{2}, & \text{if } A \text{ contains odd number of elements,} \\ +\infty, & \text{if } A \text{ is infinite.} \end{cases}$$

It is easy to see that the only measurable subsets are  $\emptyset$  and  $X$ .

**Example 11.1.4.** Let  $\mathcal{C}$  be the collection of single element subsets and two element subsets of  $X$ . Let  $\lambda(C) = 2$  if  $C$  contains single element, and  $\lambda(C) = 1$  if  $C$  contains two elements. Then  $\lambda$  induces the outer measure in Example 11.1.3, unless  $X$  contains only one element. Note that  $\lambda(C)$  for single element subset does not affect the outer measure.

**Example 11.1.5.** Let  $\mathcal{C}$  be the collection of all finite subsets of  $X$ . Let  $\lambda(C) = |C|^2$  be the square of the number of elements in  $C$ . Then the induced outer measure  $\mu^*(A) = |A|$  counts the number of elements. Moreover, every subset of  $X$  is measurable, and the measure is the counting measure. Note that if  $C$  contains more than one element, the measure  $\mu(C)$  is not  $\lambda(C)$ .

**Exercise 11.1.** In the set up of Proposition 11.1.1, prove that if a subset  $A$  is contained in a union of countably many subsets of finite outer measure, then  $A$  is contained in a union of countably many  $C_i \in \mathcal{C}$  with  $\lambda(C_i) < +\infty$ .

**Exercise 11.2.** Suppose  $\phi: X \rightarrow X$  is an invertible map, such that  $C \in \mathcal{C} \iff \phi(C) \in \mathcal{C}$ , and  $\lambda(\phi(C)) = \lambda(C)$ . Prove that the outer measure induced by  $\lambda$  satisfies  $\mu^*(\phi(A)) = \mu^*(A)$ . By Exercise 9.31, the induced measure is also invariant under  $\phi$ .

**Exercise 11.3.** Suppose  $\lambda$  is a non-negative extended valued function on a collection  $\mathcal{C}$  of subsets. Suppose  $\mu^*$  is the outer measure induced by  $\lambda$  and  $\nu^*$  is another outer measure. Then  $\nu^* \leq \mu^*$  if and only if  $\nu^*(C) \leq \lambda(C)$  for any  $C \in \mathcal{C}$ .

**Exercise 11.4.** Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra on  $X$ , and  $\lambda_1, \lambda_2$  are pre-measures on  $\mathcal{C}$ . Then  $\lambda = \lambda_1 + \lambda_2$  is also a pre-measures on  $\mathcal{C}$ . Let  $\mu_1^*, \mu_2^*, \mu^*$  be outer measures induced by  $\lambda_1, \lambda_2, \lambda$ . Let  $(X, \Sigma_1, \mu_1), (X, \Sigma_2, \mu_2), (X, \Sigma, \mu)$  be the measure spaces induced by  $\lambda_1, \lambda_2, \lambda$ .

1. Prove that  $\mu_1^* + \mu_2^* = \mu^*$ .
2. Prove that  $\Sigma_1 \cap \Sigma_2 \subset \Sigma$ , and  $\mu(A) = \mu_1(A) + \mu_2(A)$  for  $A \in \Sigma_1 \cap \Sigma_2$ .
3. Prove that if  $\mu_1(X), \mu_2(X) < +\infty$ , then  $\Sigma_1 \cap \Sigma_2 = \Sigma$ .

**Exercise 11.5.** Prove that the usual length on any of the following collections induces the Lebesgue outer measure on  $\mathbb{R}$ .

1. Bounded open intervals.
2. Bounded closed intervals.

3. Bounded intervals of the form  $[a, b)$ .
4. Bounded intervals.
5. Open intervals of length  $< 1$ .

**Exercise 11.6.** Let  $f(x)$  be a non-negative function on  $[0, +\infty)$ . Let  $\mathcal{C} = \{[a, b) : a \leq b\}$  and  $\lambda[a, b) = f(b - a)$ .

1. Prove that if  $f(x) \geq x$ ,  $f(0) = 0$ ,  $f'_+(0) = 1$ , then  $\mathcal{C}$  and  $\lambda$  induce the Lebesgue outer measure.
2. Prove that if  $f(x) \geq c$  for a constant  $c$  and all  $x > 0$ , then the only subsets measurable with respect to the outer measure induced by  $\mathcal{C}$  and  $\lambda$  are  $\emptyset$  and  $\mathbb{R}$ .

## Extension of Special Volume to Measure

Example 11.1.3 shows that, in Proposition 11.1.1, the subsets in  $\mathcal{C}$  may not be measurable with respect to the induced outer measure. In order for the subsets in  $\mathcal{C}$  to be measurable, and the measure to be  $\lambda$ , we need  $(X, \mathcal{C}, \lambda)$  to have some aspects of a measure space.

**Definition 11.1.2.** A collection  $\mathcal{C}$  of subsets is a *pre- $\sigma$ -algebra* if  $C, C' \in \mathcal{C}$  implies  $C \cap C'$  and  $C - C'$  are countable disjoint unions of subsets in  $\mathcal{C}$ . A non-negative extended valued function  $\lambda$  on  $\mathcal{C}$  is a *pre-measure* if  $C_i, \sqcup C_i \in \mathcal{C}$  implies  $\lambda(\sqcup C_i) = \sum \lambda(C_i)$ .

The terminology “pre- $\sigma$ -algebra” is not widely accepted, and “pre-measure” may have slightly different meaning in other literatures.

**Proposition 11.1.3.** Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra. Then the collection  $\mathcal{D}$  of countable disjoint unions of subsets in  $\mathcal{C}$  has the following properties.

1. If  $D, D' \in \mathcal{D}$ , then  $D \cap D' \in \mathcal{D}$ .
2. If  $D \in \mathcal{D}$  and  $C_1, \dots, C_n \in \mathcal{C}$ , then  $D - C_1 - \dots - C_n \in \mathcal{D}$ .
3. If countably many disjoint  $D_i \in \mathcal{D}$ , then  $\sqcup D_i \in \mathcal{D}$ .
4. If countably many  $C_i \in \mathcal{C}$ , then  $\cup C_i \in \mathcal{D}$ .

*Proof.* We prove the third statement first. We have  $D_i = \sqcup_j C_{ij}$ ,  $C_{ij} \in \mathcal{C}$ . Since  $D_i$  are disjoint, we know all  $C_{ij}$  are also disjoint. Then we have  $\sqcup D_i = \sqcup_{ij} C_{ij} \in \mathcal{D}$ .

For  $D, D' \in \mathcal{D}$ , we have  $D = \sqcup_i C_i$ ,  $D' = \sqcup_j C'_j$ ,  $C_i, C'_j \in \mathcal{C}$ . Since  $\mathcal{C}$  is a pre- $\sigma$ -algebra, we have  $C_i \cap C'_j \in \mathcal{D}$ . Then by the third statement, we get  $D \cap D' = \sqcup_{ij} (C_i \cap C'_j) \in \mathcal{D}$ . This proves the first statement.

For  $D \in \mathcal{D}$ , we have  $D = \sqcup_i C_i$ ,  $C_i \in \mathcal{C}$ . For  $C \in \mathcal{C}$ , since  $\mathcal{C}$  is a pre- $\sigma$ -algebra, we have  $C_i - C \in \mathcal{D}$ . Then by the third statement, we get  $D - C = \sqcup_i (C_i - C) \in \mathcal{D}$ . Now for  $C_1, \dots, C_n \in \mathcal{C}$ , by repeatedly using what we just proved, we get

$$D - C_1 \in \mathcal{D}, \quad D - C_1 - C_2 \in \mathcal{D}, \quad \dots, \quad D - C_1 - \dots - C_n \in \mathcal{D}.$$

This proves the second statement.

Suppose  $C_i \in \mathcal{C}$ . Then  $\cup_i C_i = \sqcup_i D'_i$ , with  $D'_i = C_i - C_1 - \cdots - C_{i-1}$ . We have  $D'_i \in \mathcal{D}$  by the second statement. Then by the third statement, we get  $\cup_i C_i = \sqcup_i D'_i \in \mathcal{D}$ . This proves the fourth statement.  $\square$

We may extend the pre-measure to countable disjoint unions of subsets in  $\mathcal{C}$  by

$$\lambda(D) = \sum \lambda(C_i), \quad D = \sqcup_i C_i \in \mathcal{D}, \quad C_i \in \mathcal{C}.$$

To show that the extension is well defined, we also consider  $D = \sqcup_j C'_j$ ,  $C'_j \in \mathcal{C}$ . Since  $\mathcal{C}$  is a pre- $\sigma$ -algebra, we have

$$C_i \cap C'_j = \sqcup_k C_{ijk}, \quad C_{ijk} \in \mathcal{C}.$$

Then

$$C_i = \sqcup_j (C_i \cap C'_j) = \sqcup_{jk} C_{ijk}, \quad C'_j = \sqcup_i (C_i \cap C'_j) = \sqcup_{ik} C_{ijk}.$$

By the countable additivity of  $\lambda$  on  $\mathcal{C}$ , we have

$$\sum_i \lambda(C_i) = \sum_i \sum_{jk} \lambda(C_{ijk}) = \sum_j \sum_{ik} \lambda(C_{ijk}) = \sum_j \lambda(C'_j).$$

This proves that the extension of  $\lambda$  to  $\mathcal{D}$  is well defined.

**Proposition 11.1.4.** *Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra and  $\lambda$  is a pre-measure on  $\mathcal{C}$ . Then the extension of  $\lambda$  to the collection  $\mathcal{D}$  of countable disjoint unions of subsets in  $\mathcal{C}$  has the following properties.*

1.  $\lambda(\sqcup_i D_i) = \sum \lambda(D_i)$ .
2.  $D \subset D'$  implies  $\lambda(D) \leq \lambda(D')$ .
3. The outer measure induced by  $\lambda$  is  $\mu^*(A) = \inf\{\lambda(D) : A \subset D, D \in \mathcal{D}\}$ .

*Proof.* In the first statement, let  $D_i = \sqcup_j C_{ij}$ ,  $C_{ij} \in \mathcal{C}$ . Then  $\sqcup_i D_i = \sqcup_{ij} C_{ij}$ , and

$$\lambda(\sqcup_i D_i) = \sum_{ij} \lambda(C_{ij}) = \sum_i \sum_j \lambda(C_{ij}) = \sum_i \lambda(D_i).$$

In the second statement, let  $D = \sqcup_i C_i \subset D'$ ,  $C_i \in \mathcal{C}$ ,  $D' \in \mathcal{D}$ . Then any finite partial union  $D_n = C_1 \sqcup \cdots \sqcup C_n \subset D'$ . By the second statement of Proposition 11.1.3, we have  $D' - D_n \in \mathcal{D}$ . Then by the first statement, we get

$$\lambda(D') = \lambda(D_n \sqcup (D' - D_n)) = \lambda(D_n) + \lambda(D' - D_n) \geq \lambda(D_n) = \lambda(C_1) + \cdots + \lambda(C_n).$$

Since  $n$  is arbitrary, we get

$$\lambda(D') \geq \sum_i \lambda(C_i) = \lambda(D).$$

In the third statement, we note that  $A \subset D = \sqcup_i C_i$ ,  $C_i \in \mathcal{C}$ , is a special case of the definition of the induced outer measure  $\mu^*(A)$ . Therefore  $\mu^*(A) \leq \sum_i \lambda(C_i) = \lambda(D)$ . On the other hand, for any  $\epsilon > 0$ , there is  $A \subset D = \cup_i C_i$ ,  $C_i \in \mathcal{C}$ , such that

$$\mu^*(A) + \epsilon > \sum_i \lambda(C_i).$$

We have  $D = \sqcup D'_i$ ,  $D'_i = C_i - C_1 - \cdots - C_{i-1} \in \mathcal{D}$ . By  $D'_i \subset C_i$  and the second statement, we have  $\lambda(D'_i) \leq \lambda(C_i)$ . Then by the first statement, we have

$$\mu^*(A) + \epsilon > \sum_i \lambda(C_i) \geq \sum_i \lambda(D'_i) = \lambda(D).$$

This completes the proof of the third statement.  $\square$

With the help of the extended  $\lambda$  on  $\mathcal{D}$ , we can establish the Extension Theorem. Exercise 11.8 gives an easier finite version of the theorem.

**Theorem 11.1.5 (Extension Theorem).** *Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra and  $\lambda$  is a pre-measure on  $\mathcal{C}$ . Then any subset  $C \in \mathcal{C}$  is measurable with respect to the outer measure induced by  $\lambda$ , and  $\mu(C) = \lambda(C)$ .*

*Proof.* To prove that  $C \in \mathcal{C}$  is measurable, it is sufficient to prove (9.3.2). By the third statement of Proposition 11.1.4, for any  $Y$  and  $\epsilon > 0$ , we have  $Y \subset D$ ,  $D \in \mathcal{D}$ , satisfying

$$\mu^*(Y) + \epsilon > \lambda(D).$$

By Proposition 11.1.3, we have  $Y \cap C \subset D \cap C \in \mathcal{D}$  and  $Y - C \subset D - C \in \mathcal{D}$ . Then by Proposition 11.1.4, we have

$$\lambda(D) = \lambda(D \cap C) + \lambda(D - C) \geq \mu^*(Y \cap C) + \mu^*(Y - C).$$

Combining the inequalities together, we get

$$\mu^*(Y) + \epsilon > \mu^*(Y \cap C) + \mu^*(Y - C).$$

Since  $\epsilon$  is arbitrary, we get (9.3.2).

For the equality  $\mu(C) = \lambda(C)$ , we note that  $\lambda(C) \geq \mu^*(C)$  always holds by taking  $\cup C_i$  in Proposition 11.1.1 to be  $C$ . It remains to prove  $\lambda(C) \leq \mu^*(C)$ . By the third statement in Proposition 11.1.4, this means

$$C \subset D, D \in \mathcal{D} \implies \lambda(C) \leq \lambda(D).$$

Of course this is a consequence of the second statement of Proposition 11.1.4.  $\square$

**Example 11.1.6.** We established the Lebesgue measure as an extension of the usual length of intervals. The collection  $\mathcal{C}$  of all intervals is a pre- $\sigma$ -algebra. If we can show that the usual length is a pre-measure, then the Extension Theorem gives the Lebesgue measure theory.



The key is that the usual length is countably additive. Suppose  $\langle a, b \rangle = \sqcup \langle c_i, d_i \rangle$ . Proposition 9.1.1 already implies that  $\lambda \langle a, b \rangle$  is no smaller than any partial sum of  $\sum \lambda \langle c_i, d_i \rangle$ . Therefore we have

$$\lambda \langle a, b \rangle \geq \sum \lambda \langle c_i, d_i \rangle.$$

Conversely, for any  $\epsilon > 0$ , choose  $\epsilon_i > 0$  satisfying  $\sum \epsilon_i = \epsilon$ . Then

$$[a + \epsilon, b - \epsilon] \subset \langle a, b \rangle = \sqcup \langle c_i, d_i \rangle \subset \cup (c_i - \epsilon_i, d_i + \epsilon_i).$$

By Heine-Borel theorem (Theorem 1.5.6), we have

$$[a + \epsilon, b - \epsilon] \subset (c_{i_1} - \epsilon_{i_1}, d_{i_1} + \epsilon_{i_1}) \cup \cdots \cup (c_{i_k} - \epsilon_{i_k}, d_{i_k} + \epsilon_{i_k})$$

for finitely many intervals among  $(c_i - \epsilon_i, d_i + \epsilon_i)$ . Then Proposition 9.1.1 tells us

$$\begin{aligned} \lambda \langle a, b \rangle - 2\epsilon &= \lambda[a + \epsilon, b - \epsilon] \\ &\leq \sum_{j=1}^k \lambda(c_{i_j} - \epsilon_{i_j}, d_{i_j} + \epsilon_{i_j}) = \sum_{j=1}^k (\lambda \langle c_{i_j}, d_{i_j} \rangle + 2\epsilon_{i_j}) \\ &\leq \sum \lambda \langle c_i, d_i \rangle + 2 \sum \epsilon_i = \sum \lambda \langle c_i, d_i \rangle + 2\epsilon. \end{aligned}$$

Since this holds for any  $\epsilon > 0$ , we get

$$\lambda \langle a, b \rangle \leq \sum \lambda \langle c_i, d_i \rangle.$$

**Exercise 11.7.** Suppose  $\mathcal{C}$  is a collection of subsets of  $X$ , such that for any  $C, C' \in \mathcal{C}$ , the intersection  $C \cap C'$  and the subtraction  $C - C'$  are *finite* disjoint unions of subsets in  $\mathcal{C}$ . Let  $\mathcal{D}$  be *finite* disjoint unions of subsets in  $\mathcal{C}$ . Prove that  $D, D' \in \mathcal{D}$  implies  $D \cap D', D \cup D', D - D' \in \mathcal{D}$ . Moreover, finite unions and intersections of subsets in  $\mathcal{C}$  also belong to  $\mathcal{D}$ .

**Exercise 11.8 (Carathéodory's Extension Theorem).** Suppose  $\mathcal{C}$  is a collection of subsets of  $X$ , such that for any  $C, C' \in \mathcal{C}$ , the intersection  $C \cap C'$  and the subtraction  $C - C'$  are *finite* disjoint unions of subsets in  $\mathcal{C}$ . Suppose  $\lambda$  is a non-negative extended valued function on  $\mathcal{C}$  satisfying the following conditions.

1. Additive on  $\mathcal{C}$ : If finitely many  $C_i \in \mathcal{C}$  are disjoint and  $C_1 \sqcup C_2 \sqcup \cdots \sqcup C_n \in \mathcal{C}$ , then  $\lambda(C_1 \sqcup C_2 \sqcup \cdots \sqcup C_n) = \lambda(C_1) + \lambda(C_2) + \cdots + \lambda(C_n)$ .
2. Countably subadditive on  $\mathcal{C}$ : If  $C \in \mathcal{C}$  and countably many  $C_i \in \mathcal{C}$  are disjoint, such that  $C \subset \sqcup C_i$ , then  $\lambda(C) \leq \sum \lambda(C_i)$ .

Prove that  $\lambda$  is countably additive on  $\mathcal{C}$ , and is therefore a pre-measure. In particular, the Extension Theorem (Theorem 11.1.5) can be applied to extend  $\lambda$  to a measure.

## Characterization of Measurable Subsets

In the Extension Theorem (Theorem 11.1.5), we know the subsets in the original collection  $\mathcal{C}$  are measurable. The following characterizes all measurable subsets.

**Proposition 11.1.6.** Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra and  $\lambda$  is a pre-measure on  $\mathcal{C}$ . Then a subset  $A$  with  $\mu^*(A) < +\infty$  is measurable if and only if there is a subset

$B = \cap_i (\cup_j C_{ij})$ , where the union and the intersection are countable and  $C_{ij} \in \mathcal{C}$ , such that

$$\mu(C_{ij}) < +\infty, \quad A \subset B, \quad \mu^*(B - A) = 0.$$

Moreover, it is possible to choose  $C_{ij}$  to be disjoint for the same  $i$  and distinct  $j$ , and  $\cup_j C_{ij}$  to be decreasing in  $i$ .

The conclusion remains true when  $A$  is “ $\sigma$ -finite”. See Exercise 11.10.

*Proof.* For the sufficiency part, we note that  $\mu^*(B - A) = 0$  implies the measurability of  $B - A$ . Since  $B$  is measurable by Theorem 11.1.5, we know  $A = B - (B - A)$  is measurable.

For the necessary part, we assume  $A$  is measurable, with  $\mu(A) < +\infty$ . By the third statement of Proposition 11.1.4, for any  $\epsilon > 0$ , we have  $A \subset D$ ,  $D \in \mathcal{D}$ , satisfying  $\mu(A) + \epsilon > \lambda(D)$ . By Theorem 11.1.5, the subsets in  $\mathcal{C}$  are measurable, and therefore the countable disjoint union  $D$  of subsets in  $\mathcal{C}$  is also measurable. We have  $\mu(D) = \lambda(D)$  by  $\mu = \lambda$  on  $\mathcal{C}$  and the countable additivity of  $\mu$  and  $\lambda$  on  $\mathcal{D}$  (see the first statement of Proposition 11.1.4). Then we get  $\mu(D - A) = \mu(D) - \mu(A) = \lambda(D) - \mu(A) < \epsilon$ .

Choose a sequence  $\epsilon_i \rightarrow 0$ , we get the corresponding  $D_i$  satisfying  $\mu(D_i - A) < \epsilon_i$ . By the first statement of Proposition 11.1.3, we may even replace  $D_i$  by  $D_1 \cap \cdots \cap D_i$ , so that  $D_i$  is decreasing. Then  $B = \cap_i D_i$  is of the form described in the proposition (even with  $C_{ij}$  disjoint for the same  $i$  and distinct  $j$ ). Moreover, we have  $A \subset B$  and  $\mu(B - A) \leq \mu(D_i - A) < \epsilon_i$ . Since the inequality holds for all  $i$ , we get  $\mu(B - A) = 0$ . This completes the proof of the necessity part.  $\square$

**Exercise 11.9.** Prove that in Proposition 11.1.1, for any subset  $A$ , there is a subset  $B = \cap_i (\cup_j C_{ij})$ , where the union and the intersection are countable and  $C_{ij} \in \mathcal{C}$ , such that

$$A \subset B, \quad \mu^*(A) = \mu^*(B).$$

**Exercise 11.10.** Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra and  $\lambda$  is a pre-measure on  $\mathcal{C}$ . Suppose a subset  $A$  is contained in a countable union of subsets with finite outer measure.

1. Prove that  $A$  is contained in a countable union of subsets in  $\mathcal{C}$  with finite  $\lambda$ .
2. Prove that the conclusion of Proposition 11.1.6 still holds for  $A$ .

**Exercise 11.11.** Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra and  $\lambda$  is a pre-measure on  $\mathcal{C}$ . Prove that a subset  $A$  with  $\mu^*(A) < +\infty$  is measurable if and only if for any  $\epsilon > 0$ , there is a finite union  $B = C_1 \cup \cdots \cup C_n$ ,  $C_i \in \mathcal{C}$ , such that  $\mu^*((A - B) \cup (B - A)) < \epsilon$ .

**Exercise 11.12.** Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra and  $\lambda$  is a pre-measure on  $\mathcal{C}$ . Suppose  $X$  is a union of countably many subsets in  $\mathcal{C}$  with finite  $\lambda$ . Prove that a subset  $A$  is measurable if and only if there is a subset  $B = \cup_i (\cap_j C_{ij})$ , where the union and the intersection are countable and  $C_{ij} \in \mathcal{C}$ , such that

$$A \supset B, \quad \mu^*(A - B) = 0.$$

**Exercise 11.13.** Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra and  $\lambda$  is a pre-measure on  $\mathcal{C}$ . Suppose the induced measure space  $(X, \Sigma, \mu)$  is  $\sigma$ -finite. Suppose  $f$  is a measurable function, such that  $\int_C f d\mu = 0$  for any  $C \in \mathcal{C}$ . Prove that  $f = 0$  almost everywhere. This extends Exercises 10.15 and 10.30.

**Exercise 11.14.** Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra and  $\lambda$  is a pre-measure on  $\mathcal{C}$ . Suppose  $f$  is an integrable function on the induced measure space  $(X, \Sigma, \mu)$ . Prove that for any  $\epsilon > 0$ , there are finitely many disjoint  $C_1, \dots, C_n \in \mathcal{C}$ , such that

$$\sum_{i=1}^n \left| \int_{C_i} f d\mu \right| + \epsilon > \int_X |f| d\mu.$$

This implies

$$\sup_{\text{disjoint } C_1, \dots, C_n \in \mathcal{C}} \sum_{i=1}^n \left| \int_{C_i} f d\mu \right| = \int_X |f| d\mu.$$

**Exercise 11.15.** Suppose  $\mathcal{C}$  is a pre- $\sigma$ -algebra and  $\lambda$  is a pre-measure on  $\mathcal{C}$ . Suppose  $F$  is a integrable vector valued function (see Exercises 10.57 through 10.73) on the induced measure space  $(X, \Sigma, \mu)$ . Prove that

$$\sup_{\text{disjoint } C_1, \dots, C_n \in \mathcal{C}} \sum_{i=1}^n \left\| \int_{C_i} F d\mu \right\| = \int_X \|F\| d\mu.$$

This is a more precise result than Exercise 10.71.

**Exercise 11.16.** Suppose  $f$  is a Lebesgue integrable function on  $[a, b]$ . Prove that

$$\sup_{\text{partitions of } [a, b]} \sum_{i=1}^n \left| \int_{x_{i-1}}^{x_i} f(x) dx \right| = \int_a^b |f(x)| dx.$$

Moreover, extend the result to vector valued functions.

## 11.2 Lebesgue-Stieltjes Measure

If  $\nu$  is a measure on the Borel  $\sigma$ -algebra on  $\mathbb{R}$ , then  $\alpha(x) = \nu(-\infty, x)$  is an increasing function. Strictly speaking, we need to worry about the infinity value, which happens even when  $\nu$  is the usual Lebesgue measure. Therefore we assume  $\nu$  has finite value on any bounded interval and introduce

$$\alpha_\nu(x) = \begin{cases} \nu[0, x), & \text{if } x > 0, \\ -\nu[x, 0), & \text{if } x \leq 0. \end{cases} \quad (11.2.1)$$

Then  $\alpha_\nu$  is an increasing function satisfying  $\nu[a, b] = \alpha_\nu(b) - \alpha_\nu(a)$ .

Conversely, for an increasing function  $\alpha$ , we may introduce the  $\alpha$ -length  $\lambda_\alpha[a, b] = \alpha(b) - \alpha(a)$  for the interval  $[a, b]$ . Then we may use the Extension Theorem (Theorem 11.1.5) to produce a measure  $\mu_\alpha$ .

So we expect a correspondence between measure on the Borel  $\sigma$ -algebra and increasing function. For the correspondence to become exact (say, one-to-one),

however, some conditions are needed. For example, the function  $\alpha_\nu$  must be left continuous because for any strictly increasing sequence  $b_n \rightarrow b$ , by the monotone limit property in Proposition 9.4.4, we have

$$\alpha_\nu(b) = \nu[0, b) = \nu(\cup[0, b_n)) = \lim_{n \rightarrow \infty} \nu[0, b_n) = \lim_{n \rightarrow \infty} \alpha_\nu(b_n).$$

### Lebesgue-Stieltjes Measure Induced by Increasing Function

Let  $\alpha$  be an increasing function on  $\mathbb{R}$  taking no extended value. We wish to apply the Extension Theorem (Theorem 11.1.5) to extend  $\lambda_\alpha\langle a, b \rangle = \alpha(b) - \alpha(a)$  to a measure  $\mu_\alpha$ .

Let  $\epsilon_n$  be a decreasing sequence converging to 0. By the monotone limit property in Proposition 9.4.4, the formula  $\lambda_\alpha\langle a, b \rangle = \alpha(b) - \alpha(a)$  will imply

$$\begin{aligned} \lambda_\alpha(a, b) &= \mu_\alpha(a, b) = \mu_\alpha(\cup_n \langle a + \epsilon_n, b - \epsilon_n \rangle) \\ &= \lim_{n \rightarrow \infty} \mu_\alpha\langle a + \epsilon_n, b - \epsilon_n \rangle = \lim_{n \rightarrow \infty} \lambda_\alpha\langle a + \epsilon_n, b - \epsilon_n \rangle \\ &= \lim_{n \rightarrow \infty} (\alpha(b - \epsilon_n) - \alpha(a + \epsilon_n)) = \alpha(b^-) - \alpha(a^+). \end{aligned}$$

Therefore the formula  $\lambda_\alpha\langle a, b \rangle = \alpha(b) - \alpha(a)$  is rather problematic, and should be replaced by

$$\begin{aligned} \lambda_\alpha(a, b) &= \alpha(b^-) - \alpha(a^+), & \lambda_\alpha[a, b] &= \alpha(b^+) - \alpha(a^-), \\ \lambda_\alpha(a, b] &= \alpha(b^+) - \alpha(a^+), & \lambda_\alpha[a, b) &= \alpha(b^-) - \alpha(a^-). \end{aligned}$$

**Exercise 11.17.** Justify  $\lambda_\alpha[a, b] = \alpha(b^+) - \alpha(a^-)$ .

**Exercise 11.18.** Use  $\lambda_\alpha(a, b) = \alpha(b^-) - \alpha(a^+)$  and Exercise 2.35 to justify  $\lambda_\alpha(a, b] = \alpha(b^+) - \alpha(a^+)$ .

To simplify the notation, we denote

$$\langle a^+, b^- \rangle = (a, b), \quad \langle a^-, b^+ \rangle = [a, b], \quad \langle a^+, b^+ \rangle = (a, b], \quad \langle a^-, b^- \rangle = [a, b).$$

For example, a disjoint union such as  $(a, b] = (a, c) \sqcup [c, b]$  can be written as  $\langle a^+, b^+ \rangle = \langle a^+, c^- \rangle \sqcup \langle c^-, b^+ \rangle$ , with the same symbol  $c^-$  for both intervals.

From now on, we will insist that the ends  $a, b$  in the notation  $\langle a, b \rangle$  are always be decorated with  $\pm$ . Then the four equalities defining the  $\alpha$ -length become

$$\lambda_\alpha\langle a, b \rangle = \alpha(b) - \alpha(a),$$

where we emphasize that  $a$  and  $b$  are decorated with  $\pm$ .

We remark that  $\lambda_\alpha$  depends only on the left limit functions  $\alpha(x^-)$  and the right limit function  $\alpha(x^+)$  (in fact the two half limit functions determine each other by Exercises 2.35 and 2.36), and is independent of the choice of value  $\alpha(x) \in [\alpha(x^-), \alpha(x^+)]$  at discontinuous points. Moreover, the  $\alpha$ -length of a single point is  $\lambda_\alpha\{a\} = \lambda_\alpha\langle a^-, a^+ \rangle = \alpha(a^+) - \alpha(a^-)$ .

Next we apply the Extension Theorem to  $\lambda_\alpha$ . The collection of all intervals is a pre- $\sigma$ -algebra. It remains to verify the countable additivity, which says that

$$\langle a, b \rangle = \sqcup \langle c_i, d_i \rangle \implies \lambda_\alpha \langle a, b \rangle = \sum \lambda_\alpha \langle c_i, d_i \rangle.$$

Like the proof of Proposition 9.1.1, the inclusion  $\langle c_1, d_1 \rangle \sqcup \cdots \sqcup \langle c_n, d_n \rangle \subset \langle a, b \rangle$  implies that, by rearranging the order of intervals  $\langle c_i, d_i \rangle$ , we may assume

$$a \leq c_1 \leq d_1 \leq c_2 \leq d_2 \leq \cdots \leq c_n \leq d_n \leq b.$$

Here some  $\leq$  might be  $c^- \leq c^+$ , for which we have  $\alpha(c^-) \leq \alpha(c^+)$ . Then by  $\alpha$  increasing (this is valid even with  $\pm$  decorations), we get

$$\begin{aligned} \sum_{i=1}^n \lambda_\alpha \langle c_i, d_i \rangle &= \sum_{i=1}^n (\alpha(d_i) - \alpha(c_i)) = \alpha(d_n) - \sum_{i=2}^n (\alpha(c_i) - \alpha(d_{i-1})) - \alpha(c_1) \\ &\leq \alpha(d_n) - \alpha(c_1) \leq \alpha(b) - \alpha(a) = \lambda_\alpha \langle a, b \rangle. \end{aligned}$$

Since  $n$  is arbitrary, we get  $\sum \lambda_\alpha \langle c_i, d_i \rangle \leq \lambda_\alpha \langle a, b \rangle$ .

The converse  $\sum \lambda_\alpha \langle c_i, d_i \rangle \geq \lambda_\alpha \langle a, b \rangle$  may be proved similar to the proof of Proposition 9.1.4. By Exercise 2.35,  $\alpha(x^-)$  is left continuous and  $\alpha(x^+)$  is right continuous. For any  $\epsilon > 0$ , therefore, we can find  $\delta > 0$ , such that

$$\alpha((a + \delta)^-) < \alpha(a^-) + \epsilon, \quad \alpha((b - \delta)^+) > \alpha(b^+) - \epsilon.$$

Then we try to approximate  $\langle a, b \rangle$  by closed interval  $[a', b']$  from inside

$$\langle a, b \rangle \supset [a', b'] = \begin{cases} [a, b], & \text{if } \langle a, b \rangle = \langle a^-, b^+ \rangle = [a, b], \\ [a + \delta, b], & \text{if } \langle a, b \rangle = \langle a^+, b^+ \rangle = (a, b], \\ [a, b - \delta], & \text{if } \langle a, b \rangle = \langle a^-, b^- \rangle = [a, b), \\ [a + \delta, b - \delta], & \text{if } \langle a, b \rangle = \langle a^+, b^- \rangle = (a, b). \end{cases}$$

In all four cases, we always have

$$\lambda_\alpha [a', b'] = \alpha(b'^+) - \alpha(a'^-) > \lambda_\alpha \langle a, b \rangle - 2\epsilon.$$

Similarly, we choose  $\epsilon_i > 0$  satisfying  $\sum \epsilon_i = \epsilon$ , and then find  $\delta_i > 0$ , such that

$$\alpha((c_i - \delta_i)^+) > \alpha(c_i^+) - \epsilon_i, \quad \alpha((d_i + \delta_i)^-) < \alpha(d_i^-) + \epsilon_i.$$

Then we try to approximate  $\langle c_i, d_i \rangle$  by open intervals  $(c'_i, d'_i)$  from outside

$$\langle c_i, d_i \rangle \subset (c'_i, d'_i) = \begin{cases} (c_i, d_i), & \text{if } \langle c_i, d_i \rangle = \langle c_i^+, d_i^- \rangle = (c_i, d_i), \\ (c_i - \delta_i, d_i), & \text{if } \langle c_i, d_i \rangle = \langle c_i^-, d_i^- \rangle = [c_i, d_i), \\ (c_i, d_i + \delta_i), & \text{if } \langle c_i, d_i \rangle = \langle c_i^+, d_i^+ \rangle = (c_i, d_i], \\ (c_i - \delta_i, d_i + \delta_i), & \text{if } \langle c_i, d_i \rangle = \langle c_i^-, d_i^+ \rangle = [c_i, d_i], \end{cases}$$

We always have

$$\lambda_\alpha (c'_i, d'_i) = \alpha(d_i'^-) - \alpha(c_i'^+) < \lambda_\alpha \langle c_i, d_i \rangle + 2\epsilon_i.$$

Applying Heine-Borel theorem (Theorem 1.5.6) to

$$[a', b'] \subset \langle a, b \rangle = \sqcup \langle c_i, d_i \rangle \subset \cup (c'_i, d'_i),$$

we have

$$[a', b'] \subset (c'_{i_1}, d'_{i_1}) \cup \cdots \cup (c'_{i_k}, d'_{i_k})$$

for finitely many intervals among  $(c'_i, d'_i)$ . By rearranging the intervals if necessary, we may further assume that this implies

$$a' > c'_{i_1}, d'_{i_1} > c'_{i_2}, d'_{i_2} > c'_{i_3}, \dots, d'_{i_{k-1}} > c'_{i_k}, d'_{i_k} > b'.$$

Then we get

$$\begin{aligned} \lambda_\alpha \langle a, b \rangle - 2\epsilon &< \alpha(b'^+) - \alpha(a'^-) \\ &\leq \alpha(d'_{i_k}) - \alpha(c'_{i_k}) + \alpha(d'_{i_{k-1}}) - \alpha(c'_{i_{k-1}}) + \cdots + \alpha(d'_{i_1}) - \alpha(c'_{i_1}) \\ &\leq \sum_{j=1}^k (\lambda_\alpha \langle c_{i_j}, d_{i_j} \rangle + 2\epsilon_j) \leq \sum_{i=1}^{\infty} \lambda_\alpha \langle c_i, d_i \rangle + 2\epsilon. \end{aligned}$$

Since  $\epsilon > 0$  is arbitrary, we get  $\lambda_\alpha \langle a, b \rangle \leq \sum \lambda_\alpha \langle c_i, d_i \rangle$ .

We verified the condition that  $\lambda_\alpha$  is a pre-measure. By the Extension Theorem, therefore,  $\lambda_\alpha$  can be extended to a measure  $\mu_\alpha$  on a  $\sigma$ -algebra  $\Sigma_\alpha$  that contains all intervals. We call  $\mu_\alpha$  the *Lebesgue-Stieltjes measure* induced by  $\alpha$ . The  $\sigma$ -algebra  $\Sigma_\alpha$  contains all Borel sets, and may be different for different  $\alpha$ .

**Exercise 11.19.** Prove that the  $\alpha$ -length on any of the following collections induces the Lebesgue-Stieltjes outer measure on  $\mathbb{R}$ .

1. Bounded open intervals.
2. Bounded closed intervals.
3. Bounded intervals of the form  $[a, b)$ .
4. Bounded intervals.
5. Open intervals of length  $< 1$ .

This extends Exercise 11.5.

**Exercise 11.20.** Suppose  $\alpha, \beta$  are increasing functions. Prove that a subset is Lebesgue-Stieltjes measurable with respect to  $\alpha + \beta$  if and only if it is Lebesgue-Stieltjes measurable with respect to  $\alpha$  and  $\beta$ , and  $\mu_{\alpha+\beta}(A) = \mu_\alpha(A) + \mu_\beta(A)$ .

## Lebesgue-Stieltjes Measurability

**Proposition 11.2.1.** *The following are equivalent to the Lebesgue-Stieltjes measurability of a subset  $A$  with respect to an increasing function  $\alpha$ .*

1. For any  $\epsilon > 0$ , there is an open subset  $U$  and a closed subset  $C$ , such that  $C \subset A \subset U$  and  $\mu_\alpha(U - C) < \epsilon$ .

2. There is a countable intersection  $D$  of open subsets (called  $G_\delta$ -set) and a countable union  $S$  of closed subsets (called  $F_\sigma$ -set), such that  $S \subset A \subset D$ , and  $\mu_\alpha(D - S) = 0$ .

Moreover, in case  $\mu_\alpha^*(A)$  is finite, we may take  $C$  in the first statement to be compact.

*Proof.* Suppose  $A$  is Lebesgue-Stieltjes measurable. For any  $\epsilon > 0$ , there is a countable union  $U = \cup_i \langle c_i, d_i \rangle$  ( $c_i, d_i$  are decorated with  $\pm$ ) of intervals, such that

$$A \subset U, \quad \sum_i \lambda_\alpha \langle c_i, d_i \rangle < \mu_\alpha^*(A) + \epsilon = \mu_\alpha(A) + \epsilon.$$

As argued earlier, we can enlarge  $\langle c_i, d_i \rangle$  to an open interval  $(c'_i, d'_i)$  ( $c'_i, d'_i$  are not decorated), such that the increase from  $\lambda_\alpha \langle c_i, d_i \rangle$  to  $\lambda_\alpha(c'_i, d'_i) = \mu_\alpha(c'_i, d'_i)$  is as tiny as we want. Therefore we can still have

$$A \subset U = \cup_i (c'_i, d'_i), \quad \sum_i \mu_\alpha(c'_i, d'_i) < \mu_\alpha(A) + \epsilon.$$

Then  $U$  is open and

$$\mu_\alpha(A) \leq \mu_\alpha(U) \leq \sum_i \mu_\alpha(c'_i, d'_i) < \mu_\alpha(A) + \epsilon.$$

If  $\mu_\alpha(A)$  is finite, then we get  $\mu_\alpha(U - A) = \mu_\alpha(U) - \mu_\alpha(A) < \epsilon$ . This shows that any Lebesgue-Stieltjes measurable subset of finite measure is approximated by an open subset from outside. In general, we have  $A = \cup_{j=1}^\infty A_j$ ,  $A_j = A \cap [-j, j]$ , and we have open  $U_j \supset A_j$  satisfying  $\mu_\alpha(U_j - A_j) < \frac{\epsilon}{2^j}$ . Then we get open  $U = \cup U_j \supset A$  satisfying  $U - A = \cup(U_j - A) \subset \cup(U_j - A_j)$  and

$$\mu_\alpha(U - A) \leq \sum \mu_\alpha(U_j - A_j) < \sum \frac{\epsilon}{2^j} = \epsilon.$$

So any Lebesgue-Stieltjes measurable subset is approximated by an open subset from outside. By taking the complement, any Lebesgue-Stieltjes measurable subset is also approximated by a closed subset from inside. Specifically, the complement  $\mathbb{R} - A$  is also measurable. So for any  $\epsilon > 0$ , there is open  $V \supset \mathbb{R} - A$ , such that  $\mu_\alpha(V - (\mathbb{R} - A)) < \epsilon$ . Then  $C = \mathbb{R} - V$  is a closed subset contained in  $A$ . Moreover, we have  $V - (\mathbb{R} - A) = A - C$ , so that  $\mu_\alpha(A - C) < \epsilon$ .

Combining the outer and inner approximations, we get  $C \subset A \subset U$  satisfying  $\mu_\alpha(U - A) \leq \mu_\alpha(U - A) + \mu_\alpha(A - C) < 2\epsilon$ . This proves that the Lebesgue-Stieltjes measurability implies the first statement.

In case  $\mu_\alpha(A) < +\infty$ , by the monotone limit property in Proposition 9.4.4, we have  $\lim_{r \rightarrow +\infty} \mu_\alpha(A - C \cap [-r, r]) = \mu_\alpha(A - C) < \epsilon$ . Therefore  $\mu_\alpha(A - C \cap [-r, r]) < \epsilon$  for sufficiently big  $r$ . By replacing  $C$  with  $C \cap [-r, r]$ , we may assume that the inner approximation  $C$  is compact.

Now we assume the first statement. Then we have open  $U_j$  and closed  $C_j$ , such that

$$C_j \subset A \subset U_j, \quad \lim \mu(U_j - C_j) = 0.$$

The intersection  $D = \cap U_j$  is a  $\delta$ -set, and the union  $S = \cup C_j$  is a  $\sigma$ -set. Moreover, we have  $S \subset A \subset D$  and by  $D - S \subset U_j - C_j$ , so that

$$\mu_\alpha(D - S) \leq \mu_\alpha(U_j - C_j).$$

Then  $\lim \mu_\alpha(U_j - C_j) = 0$  implies  $\mu_\alpha(D - S) = 0$ . This proves the second statement.

Finally, assume the second statement. Then  $A - S \subset D - S$ ,  $\mu_\alpha(D - S) = 0$ , and the completeness of  $\mu_\alpha$  implies that  $A - S$  is measurable. Since we already know that the  $\sigma$ -set  $S$  is measurable, we conclude that  $A = S \cup (A - S)$  is Lebesgue-Stieltjes measurable.  $\square$

### Regular Measure on $\mathbb{R}$

The following result fulfills the early promise of one-to-one correspondence between measures on  $\mathbb{R}$  and increasing functions on  $\mathbb{R}$ , under suitable conditions.

**Theorem 11.2.2.** *There is a one-to-one correspondence between the following.*

1. *Equivalent classes of increasing functions  $\alpha$  on  $\mathbb{R}$ . Two increasing functions  $\alpha$  and  $\beta$  are equivalent if there is a constant  $C$ , such that  $\alpha(x) = \beta(x) + C$  whenever both  $\alpha$  and  $\beta$  are continuous at  $x$ .*
2. *Measures  $\nu$  on the Borel  $\sigma$ -algebra on  $\mathbb{R}$ , such that  $\nu(I) < +\infty$  for any bounded interval  $I$ , and for any Borel set  $A$  and  $\epsilon > 0$ , there is an open subset  $U \supset A$  satisfying  $\nu(U - A) < \epsilon$ .*

The assumption  $\nu(I) < +\infty$  for any bounded interval  $I$  allows us to take the complement approximation, and implies that the measure  $\nu$  of any Borel subset can also be approximated by closed subsets from inside. The measures that can be approximated by open subsets from outside as well as by closed subsets from inside are called *regular*.

*Proof.* For  $\alpha$  in the first class, by Proposition 11.2.1, the Lebesgue-Stieltjes measure  $\mu_\alpha$  is in the second class. We need to show that  $\alpha \mapsto \mu_\alpha$  is well defined. By Exercise 2.38, the discontinuity points of an increasing function must be countable. Therefore  $\alpha(x) = \beta(x) + C$  for all but countably many  $x$ . Then any  $x$  is the limit of a strictly decreasing sequence  $x_n$  satisfying  $\alpha(x_n) = \beta(x_n) + C$ . Taking the limit of the equality, we get  $\alpha(x^+) = \beta(x^+) + C$ . Similar argument gives  $\alpha(x^-) = \beta(x^-) + C$ . Therefore the  $\alpha$ -length and the  $\beta$ -length are the same, and they extend to the same Lebesgue-Stieltjes measure.

The correspondence from the second class to the first is given by  $\nu \mapsto \alpha_\nu$  defined in (11.2.1). We need to show that the two correspondences are inverse to each other.

So for an increasing  $\alpha$ , we apply (11.2.1) to the Lebesgue-Stieltjes measure  $\mu_\alpha$  to get

$$\beta(x) = \begin{cases} \mu_\alpha[0, x), & \text{if } x > 0 \\ -\mu_\alpha[x, 0), & \text{if } x \leq 0 \end{cases} = \alpha(x^-) - \alpha(0^-).$$



Let  $C = -\alpha(0^-)$ . Then  $\beta(x) = \alpha(x) + C$  whenever  $\alpha$  is left continuous at  $x$ . Therefore  $\alpha$  and  $\beta$  are equivalent.

On the other hand, for any measure  $\nu$  satisfying the condition of the proposition, we introduce  $\alpha = \alpha_\nu$  by (11.2.1). Then  $\alpha$  is left continuous and satisfies  $\nu[a, b) = \alpha(b) - \alpha(a) = \alpha(b^-) - \alpha(a^-) = \mu_\alpha[a, b)$ . Since any open interval is the union of an increasing sequence of intervals of the form  $[a, b)$ , by the monotone limit property in Proposition 9.4.4, we get  $\nu(a, b) = \mu_\alpha(a, b)$ . Then the countable additivity implies that  $\nu(U) = \mu_\alpha(U)$  for any open subset  $U$ . Now for any Borel set  $A$  contained in a bounded interval, the condition on  $A$  and the application of Proposition 11.2.1 to  $\mu_\alpha$  imply that, for any  $\epsilon > 0$ , there are open  $U_1, U_2 \supset A$ , such that

$$\nu(U_1 - A) < \epsilon, \quad \mu_\alpha(U_2 - A) < \epsilon.$$

The intersection  $U = U_1 \cap U_2 \supset A$  is still open, and satisfies

$$\begin{aligned} 0 &\leq \nu(U) - \nu(A) = \nu(U - A) \leq \nu(U_1 - A) < \epsilon, \\ 0 &\leq \mu_\alpha(U) - \mu_\alpha(A) = \mu_\alpha(U - A) \leq \mu_\alpha(U_2 - A) < \epsilon. \end{aligned}$$

By  $\nu(U) = \mu_\alpha(U)$ , this implies  $|\nu(A) - \mu_\alpha(A)| < \epsilon$ . Since this is true for any  $\epsilon$ , we conclude that  $\nu(A) = \mu_\alpha(A)$  for any Borel set contained in a bounded interval. For general Borel set  $A$ , we then have

$$\nu(A) = \lim_{r \rightarrow +\infty} \nu(A \cap [-r, r]) = \lim_{r \rightarrow +\infty} \mu_\alpha(A \cap [-r, r]) = \mu_\alpha(A).$$

This completes the proof that the correspondences are inverse to each other.  $\square$

**Exercise 11.21.** Prove the following are equivalent for increasing functions  $\alpha$  and  $\beta$ .

1. If  $\alpha$  and  $\beta$  are continuous at  $x$ , then  $\alpha(x) = \beta(x)$ .
2.  $\alpha(x^+) = \beta(x^+)$ .
3.  $\alpha(x^-) = \beta(x^-)$ .
4.  $\beta(x^-) \leq \alpha(x) \leq \beta(x^+)$ .

**Exercise 11.22.** Prove that, up to adding constants, each class of increasing functions in Theorem 11.2.2 contains a unique left continuous function.

## 11.3 Product Measure

Let  $(X, \Sigma_X, \mu)$  and  $(Y, \Sigma_Y, \nu)$  be measure spaces. A *measurable rectangle* is a subset  $A \times B$  of  $X \times Y$  with  $A \in \Sigma_X$  and  $B \in \Sigma_Y$ . Define the measure of the measurable rectangle to be

$$\mu \times \nu(A \times B) = \mu(A)\nu(B).$$

The formula also means that, if  $\mu(A) = 0$ , then  $\mu \times \nu(A \times B) = 0$  no matter whether  $\nu(B)$  is finite or  $+\infty$ . In case  $\nu(B) = 0$ , we also take  $\mu \times \nu(A \times B) = 0$ .

By Proposition 11.1.1, this gives an outer measure for subsets  $E \subset X \times Y$

$$(\mu \times \nu)^*(E) = \inf \left\{ \sum \mu(A_i)\nu(B_i) : E \subset \bigcup A_i \times B_i, A_i \in \Sigma_X, B_i \in \Sigma_Y \right\}.$$

**Definition 11.3.1.** Given measure spaces  $(X, \Sigma_X, \mu)$  and  $(Y, \Sigma_Y, \nu)$ , the *complete product measure*  $(X \times Y, \overline{\Sigma_X \times \Sigma_Y}, \overline{\mu \times \nu})$  is the measure space induced by the outer measure  $(\mu \times \nu)^*$ .

**Proposition 11.3.2.** *Measurable rectangles are measurable in the complete product measure space, and the complete product measure extends the measure of measurable rectangles.*

*Proof.* We simply need to verify the conditions of the Extension Theorem (Theorem 11.1.5). For measurable triangles  $A \times B$  and  $A' \times B'$ , we have

$$\begin{aligned} A \times B \cap A' \times B' &= (A \cap A') \times (B \cap B'), \\ A \times B - A' \times B' &= A \times (B - B') \sqcup (A - A') \times (B \cap B'). \end{aligned}$$

This shows that the collection of measurable triangles is a pre- $\sigma$ -algebra.

Next we verify the countable additivity. For a countable disjoint union  $A \times B = \sqcup A_i \times B_i$ ,  $A_i \in \Sigma_X$ ,  $B_i \in \Sigma_Y$ , we have

$$\chi_A(x)\chi_B(y) = \sum \chi_{A_i}(x)\chi_{B_i}(y).$$

For fixed  $y$ , the right side is a non-negative series of measurable functions on  $X$ . By the Monotone Convergence Theorem (Theorem 10.4.2), we have

$$\begin{aligned} \mu(A)\chi_B(y) &= \int_X \chi_A(x)\chi_B(y)d\mu = \int_X \sum \chi_{A_i}(x)\chi_{B_i}(y)d\mu \\ &= \sum \int_X \chi_{A_i}(x)\chi_{B_i}(y)d\mu = \sum \mu(A_i)\chi_{B_i}(y). \end{aligned}$$

The right side is again a non-negative series of measurable functions on  $X$ , and applying the Monotone Convergence Theorem again gives

$$\mu(A)\nu(B) = \int_Y \mu(A)\chi_B(y)d\nu = \sum \int_Y \mu(A_i)\chi_{B_i}(y)d\nu = \sum \mu(A_i)\nu(B_i).$$

This verifies the countable additivity. □

## Fubini Theorem

**Proposition 11.3.3.** *Suppose  $E \in \overline{\Sigma_X \times \Sigma_Y}$  satisfies  $\overline{\mu \times \nu}(E) < +\infty$ . Then for almost all  $x \in X$ , the section of  $E$  at  $x \in X$*

$$E_x = \{y \in Y : (x, y) \in E\}$$

*is almost the same as a measurable subset in  $Y$ , and  $\nu(E_x) < +\infty$ . Moreover,  $\nu(E_x)$  is equal to a measurable function on  $X$  almost everywhere, and*

$$\overline{\mu \times \nu}(E) = \int_X \nu(E_x)d\mu. \quad (11.3.1)$$

Strictly speaking,  $\nu(E_x)$  may not be defined because  $E_x$  may not be measurable. However, if  $E_x$  is almost equal to a measurable subset  $B \in \Sigma_Y$ , then Propositions 9.4.6 and 9.4.8 shows that  $\nu(B)$  is independent of the choice of  $B$ , and can be defined as  $\nu(E_x)$ .

*Proof.* First, we consider countable disjoint union of measurable rectangles

$$E = \sqcup A_j \times B_j, \quad A_j \in \Sigma_X, \quad B_j \in \Sigma_Y, \quad \sum \mu(A_j)\nu(B_j) < +\infty. \quad (11.3.2)$$

For any fixed  $x \in X$ , the section  $E_x = \sqcup_{x \in A_j} B_j$  is always measurable, and

$$\nu(E_x) = \sum_{x \in A_j} \nu(B_j) = \sum \chi_{A_j}(x)\nu(B_j).$$

Therefore  $\nu(E_x)$  is a measurable function on  $X$ , and by the Monotone Convergence Theorem (Theorem 10.4.2), we have

$$\begin{aligned} \int_X \nu(E_x) d\mu &= \sum \int_X \chi_{A_j}(x)\nu(B_j) d\mu = \sum \mu(A_j)\nu(B_j) \\ &= \sum \overline{\mu \times \nu}(A_j \times B_j) = \overline{\mu \times \nu}(E). \end{aligned}$$

Moreover,  $\overline{\mu \times \nu}(E) < +\infty$  implies that  $\nu(E_x) < +\infty$  for almost all  $x$ .

Second, we consider countable intersection of the first case

$$E = \cap E_i, \quad \text{each } E_i \text{ is of the form (11.3.2)}. \quad (11.3.3)$$

By the first statement in Proposition 11.1.3,  $E_1 \cap \dots \cap E_i$  is also of the form (11.3.2). Therefore we may further assume that  $E_i \supset E_{i+1}$ . From the first case, we know  $(E_i)_x$  is always measurable and has finite measure for almost all  $x$ . Therefore  $E_x = \cap (E_i)_x$  is always measurable and has finite measure for almost all  $x$ . Then by  $E_i \supset E_{i+1}$  implying  $(E_i)_x \supset (E_{i+1})_x$ , we may apply the monotone limit property in Proposition 9.4.4 to the decreasing intersection  $E_x = \cap (E_i)_x$  and get

$$\nu(E_x) = \lim \nu((E_i)_x) \text{ for almost all } x.$$

From the first case, we know each  $\nu((E_i)_x)$  is a measurable function on  $X$ . Therefore the (almost) limit function  $\nu(E_x)$  is equal to a measurable function almost everywhere, and we may apply the Monotone Convergence Theorem to get

$$\int_X \nu(E_x) d\mu = \lim \int_X \nu((E_i)_x) d\mu = \lim \overline{\mu \times \nu}(E_i),$$

where the second equality has been proved for  $E_i$  of the form (11.3.2). Then we may apply the monotone limit property in Proposition 9.4.4 again to the decreasing intersection  $E = \cap E_i$  and get

$$\overline{\mu \times \nu}(E) = \lim \overline{\mu \times \nu}(E_i).$$

This proves the equality (11.3.1) for  $E = \cap E_i$ .

Finally, we consider general  $E \in \overline{\Sigma_X \times \Sigma_Y}$  satisfying  $\overline{\mu \times \nu}(E) < +\infty$ . By Proposition 11.1.6, we have  $E = G - H$  for a subset  $G$  of the form (11.3.3), and another subset  $H$  satisfying  $\overline{\mu \times \nu}(H) = 0$ . Moreover, if we apply Proposition 11.1.6 to  $H$ , then we get  $H \subset F$  for a subset  $F$  of the form (11.3.3) satisfying  $\overline{\mu \times \nu}(F) = 0$ .

From the second case, we know that the sections  $G_x$  and  $F_x$  are always measurable, and the equality (11.3.1) holds for  $G$  and  $H$ . Note that the equality (11.3.1) for  $H$  means

$$\int_X \nu(F_x) d\mu = \overline{\mu \times \nu}(F) = 0.$$

This implies  $\nu(F_x) = 0$  for almost all  $x$ . Then

$$E_x = G_x - H_x, \quad H_x \subset F_x$$

implies that, for almost all  $x$ ,  $E_x$  is almost the same as the measurable subset  $G_x$ . Therefore we find that  $\nu(E_x) = \nu(G_x)$  for almost all  $x$ . Since  $G$  is of the form (11.3.3), the proposition is proved for  $G$ , and we get

$$\begin{aligned} \int_X \nu(E_x) d\mu &= \int_X \nu(G_x) d\mu && (\nu(E_x) = \nu(G_x) \text{ for almost all } x) \\ &= \overline{\mu \times \nu}(G) && ((11.3.1) \text{ proved for } G) \\ &= \overline{\mu \times \nu}(E). && (\overline{\mu \times \nu}(H) = 0) \end{aligned}$$

This proves (11.3.1) for general  $E$ . Moreover,  $\overline{\mu \times \nu}(E) < +\infty$  implies that  $\nu(E_x) < +\infty$  for almost all  $x$ .  $\square$

The computation of the product measure in Proposition 11.3.3 leads to the computation of the integration with respect to the product measure.

**Theorem 11.3.4 (Fubini Theorem).** *Suppose  $f(x, y)$  is integrable with respect to the complete product measure. Then for almost all  $x$ , the section  $f_x(y) = f(x, y)$  is integrable on  $Y$ . Moreover,  $\int_Y f_x d\nu$  is integrable on  $X$ , and*

$$\int_{X \times Y} f d\overline{\mu \times \nu} = \int_X \left( \int_Y f_x d\nu \right) d\mu.$$

*Proof.* We first assume that  $f$  is measurable. By Proposition 10.3.5, it is sufficient to prove for the case that  $f$  is non-negative. By Lemma 10.4.1, there is an increasing sequence of measurable simple functions  $\phi_n$  of the form (12.4.1), such that  $\lim \phi_n = f$ , and

$$\int_{X \times Y} f d\overline{\mu \times \nu} = \lim \int_{X \times Y} \phi_n d\overline{\mu \times \nu}.$$

Since the measurable subsets in  $\phi_n$  have finite measure, by Proposition 11.3.3, Fubini Theorem holds for the characteristic functions of these measurable subsets and their finite linear combinations  $\phi_n$ . In other words, the sections  $(\phi_n)_x$  are

measurable functions of  $y$  for almost all  $x$ , and  $\psi_n(x) = \int_Y (\phi_n)_x d\nu$  is equal to a measurable function almost everywhere, and

$$\int_{X \times Y} \phi_n d\mu \times \nu = \int_X \left( \int_Y (\phi_n)_x d\nu \right) d\mu = \int_X \psi_n d\mu.$$

Since  $\phi_n$  is increasing,  $\psi_n$  is also increasing. We may apply the Monotone Convergence Theorem (Theorem 10.4.2) to the right and get

$$\int_{X \times Y} f d\mu \times \nu = \lim \int_{X \times Y} \phi_n d\mu \times \nu = \lim \int_X \psi_n d\mu = \int_X \lim \psi_n d\mu.$$

On the other hand, for each  $x$ , the section  $(\phi_n)_x$  is an increasing sequence of simple functions on  $Y$  converging to the section  $f_x$ . Therefore  $f_x$  is measurable for almost all  $x$ , and by the Monotone Convergence Theorem,

$$\lim \psi_n(x) = \lim \int_Y (\phi_n)_x d\nu = \int_Y f_x d\nu \text{ for almost all } x.$$

Combined with the equality above, we get

$$\int_{X \times Y} f d\mu \times \nu = \int_X \left( \int_Y f_x d\nu \right) d\mu.$$

For the general case,  $f$  is equal to a measurable function  $g$  outside a subset  $F$  of measure zero. Since we have proved the proposition for  $g$ , we have

$$\int_{X \times Y} f d\mu \times \nu = \int_{X \times Y} g d\mu \times \nu = \int_X \left( \int_Y g_x d\nu \right) d\mu.$$

To show that the right side is equal to  $\int_X \left( \int_Y f_x d\nu \right) d\mu$ , we use Proposition 11.3.3 (especially the last part of the proof) to get  $\mu(F_x) = 0$  for almost all  $x$ . Since  $f_x = g_x$  outside  $F_x$ , for almost all  $x \in X$ , we get  $f_x = g_x$  almost everywhere on  $Y$ . This implies that  $\int_Y g_x d\nu = \int_Y f_x d\nu$  for almost all  $x$ .  $\square$

### Product $\sigma$ -Algebra

The complete product measure is always complete, despite the measures on  $X$  and  $Y$  may not be complete. This is the cause of the “messy part” in Proposition 11.3.3 and Theorem 11.3.4, that many things are only almost true. The problem can be addressed by introducing cleaner  $\sigma$ -algebra for the product measure.

**Definition 11.3.5.** Given  $\sigma$ -algebras  $\Sigma_X$  and  $\Sigma_Y$  on  $X$  and  $Y$ , the *product  $\sigma$ -algebra*  $\Sigma_X \times \Sigma_Y$  is the smallest  $\sigma$ -algebra (see Exercise 9.4.3) on  $X \times Y$  that contains all the measurable rectangles. Moreover, the *product measure*  $(X \times Y, \Sigma_X \times \Sigma_Y, \mu \times \nu)$  is the restriction of the complete product measure to the product  $\sigma$ -algebra.

**Proposition 11.3.6.** *If  $E$  and  $f$  are measurable with respect to  $\Sigma_X \times \Sigma_Y$ , then the sections  $E_x$  and  $f_x$  are measurable with respect to  $\Sigma_Y$ .*

*Proof.* It is easy to verify that collection

$$\Sigma = \{E \in X \times Y : E_x \in \Sigma_Y \text{ for all } x \in X\}$$

is a  $\sigma$ -algebra, and contains all measurable rectangles. Therefore  $\Sigma_X \times \Sigma_Y \subset \Sigma$ . This means that if  $E \in \Sigma_X \times \Sigma_Y$ , then  $E_x \in \Sigma_Y$  for any  $x \in X$ .

The similar statement for the function  $f(x, y)$  follows from

$$f_x^{-1}(a, b) = [f^{-1}(a, b)]_x. \quad \square$$

**Proposition 11.3.7.** *Suppose  $\nu$  is a  $\sigma$ -finite measure and  $E \in \Sigma_X \times \Sigma_Y$ . Then  $\nu(E_x)$  is a measurable function on  $X$ , and*

$$\mu \times \nu(E) = \int_X \nu(E_x) d\mu.$$

*Proof.* Assume  $\nu(Y)$  is finite. We put what we wish to be true together and define  $\mathcal{M}$  to be the collection of  $E \subset X \times Y$  satisfying the following.

1.  $E_x$  is measurable for each  $x \in X$ .
2.  $\nu(E_x)$  is a measurable function on  $X$ .

Since we have not yet assumed  $E$  to be measurable, we cannot yet require the formula for  $\mu \times \nu(E)$ .

It is likely that  $\mathcal{M}$  is not a  $\sigma$ -algebra. On the other hand, we are able to show that  $\mathcal{M}$  is a *monotone class* in the sense that the following hold.

- If  $E_i \in \mathcal{M}$  satisfy  $E_i \subset E_{i+1}$ , then  $\cup E_i \in \mathcal{M}$ .
- If  $E_i \in \mathcal{M}$  satisfy  $E_i \supset E_{i+1}$ , then  $\cap E_i \in \mathcal{M}$ .

For increasing  $E_i \in \mathcal{M}$ , we have  $(\cup E_i)_x = \cup (E_i)_x$ . Since  $(E_i)_x$  is measurable for all  $i$  and  $x$ , we find  $(\cup E_i)_x$  to be measurable for all  $x$ . Moreover, by the monotone limit property in Proposition 9.4.4, we have  $\nu((\cup E_i)_x) = \lim \nu((E_i)_x)$ . Then the measurability of the functions  $\nu((E_i)_x)$  implies the measurability of the function  $\nu((\cup E_i)_x)$ . This verifies the first property of the monotone class. The second properties can be verified similarly, where  $\nu(Y) < +\infty$  is used for  $\nu((\cap E_i)_x) = \lim \nu((E_i)_x)$  for decreasing  $E_i$ .

Let

$$\mathcal{A} = \{\sqcup_{i=1}^n A_i \times B_i : A_i \in \Sigma_X, B_i \in \Sigma_Y, n \in \mathbb{N}\}$$

be the collection of finite disjoint unions of measurable rectangles. It is easy to see that  $\mathcal{M} \supset \mathcal{A}$ . Moreover,  $\mathcal{A}$  is an *algebra* in the sense that the whole space belongs to  $\mathcal{A}$  and

$$E, F \in \mathcal{A} \implies E \cup F, E \cap F, E - F \in \mathcal{A}.$$

It is easy to show that the intersection of monotone classes is still a monotone class. Therefore it makes sense to introduce the smallest monotone class  $\mathcal{N}$  containing  $\mathcal{A}$ . Since  $\mathcal{M}$  is also a monotone class containing  $\mathcal{A}$ , we get  $\mathcal{M} \supset \mathcal{N}$ . Our goal is to use the minimality of  $\mathcal{N}$  to show that  $\mathcal{N}$  is also a  $\sigma$ -algebra. Then  $\mathcal{N} \supset \mathcal{A}$  implies  $\mathcal{N} \supset \Sigma_X \times \Sigma_Y$ , and we conclude  $\mathcal{M} \supset \Sigma_X \times \Sigma_Y$ . This means that any  $E \in \Sigma_X \times \Sigma_Y$  has the two defining properties of  $\mathcal{M}$ .

By Exercise 9.34, a monotone class is a  $\sigma$ -algebra if and only if it is an algebra. To prove that  $\mathcal{N}$  is a  $\sigma$ -algebra, therefore, it is sufficient to prove that  $\mathcal{N}$  is an algebra. For any  $E \in \mathcal{N}$ , we introduce

$$\mathcal{N}_E = \{F \in \mathcal{N} : F - E, E - F, E \cup F, E \cap F \in \mathcal{N} \text{ for all } E \in \mathcal{N}\}.$$

The problem then becomes  $\mathcal{N}_E \supset \mathcal{N}$ . We will prove the containment (actually equality) by using the minimality of  $\mathcal{N}$  and the obvious symmetry in the definition of  $\mathcal{N}_E$

$$F \in \mathcal{N}_E \iff E \in \mathcal{N}_F.$$

The fact that  $\mathcal{A}$  is an algebra means

$$E, F \in \mathcal{A} \implies E \in \mathcal{N}_F.$$

In other words, we have  $\mathcal{N}_F \supset \mathcal{A}$  for any  $F \in \mathcal{A}$ . It is also easy to verify that  $\mathcal{N}_F$  is a monotone class. Therefore we have  $\mathcal{N}_F \supset \mathcal{N}$  for any  $F \in \mathcal{A}$ , and we get

$$\begin{aligned} E \in \mathcal{N}, F \in \mathcal{A} &\implies E \in \mathcal{N}_F && (\mathcal{N}_F \supset \mathcal{N} \text{ for any } F \in \mathcal{A}) \\ &\implies F \in \mathcal{N}_E. && (\text{symmetry in the definition of } \mathcal{N}_F) \end{aligned}$$

This means that  $\mathcal{N}_E \supset \mathcal{A}$  for any  $E \in \mathcal{N}$ . Since  $\mathcal{N}_E$  is a monotone class, we get  $\mathcal{N}_E \supset \mathcal{N}$  for any  $E \in \mathcal{N}$ . As explained earlier, this implies  $\mathcal{M} \supset \Sigma_X \times \Sigma_Y$ .

Now we turn to the equality

$$\mu \times \nu(E) = \int_X \nu(E_x) d\mu.$$

By the Fubini Theorem (Theorem 11.3.4), we already have the equality for  $\Sigma_X \times \Sigma_Y$  under the condition  $\mu \times \nu(E) < +\infty$ . Since the condition is not assumed in the current proposition, it is worthwhile to state the equality again.

To prove the equality for  $\mu \times \nu(E)$ , we may add  $E \in \Sigma_X \times \Sigma_Y$  and the equality to the definition of  $\mathcal{M}$ . The monotone property can be verified by using the Monotone Convergence Theorem (Theorem 10.4.2), and the assumption  $\nu(Y) < +\infty$  is again needed for decreasing  $E_i$ . The rest of the argument is the same. In fact, since we added  $E \in \Sigma_X \times \Sigma_Y$  to the definition of  $\mathcal{M}$ , the new  $\mathcal{M}$  is equal to  $\Sigma_X \times \Sigma_Y$  at the end.

Finally, the case that  $\nu$  is  $\sigma$ -finite can be obtained from the finite case, by writing  $E = \cup(E \cap X \times Y_i)$ , where  $Y_i \in \Sigma_Y$  is an increasing sequence such that  $\nu(Y_i)$  are finite and  $Y = \cup Y_i$ .  $\square$

**Exercise 11.23.** Prove that  $(X \times Y, \overline{\Sigma_X \times \Sigma_Y}, \overline{\mu \times \nu})$  is the completion of  $(X \times Y, \Sigma_X \times \Sigma_Y, \mu \times \nu)$ .

**Exercise 11.24.** Let  $f \geq 0$  be a function on the measure space  $(X, \Sigma)$ . Let  $X \times \mathbb{R}$  have the product  $\sigma$ -algebra of  $\Sigma$  with the Borel  $\sigma$ -algebra. Prove that if  $f$  is measurable, then

$$G(f) = \{(x, y) : x \in X, 0 \leq y < f(x)\} \subset X \times \mathbb{R}$$

is measurable. Moreover, if  $f$  is integrable, then  $\mu(G(f)) = \int f d\mu$ . Does the measurability of  $G(f)$  imply the measurability of  $f$ ?

## 11.4 Lebesgue Measure on $\mathbb{R}^n$

The Lebesgue measure on the real line can be extended to the Euclidean space  $\mathbb{R}^n$ , by requiring rectangles to have the expected volume. For example, if we take all the rectangles  $A_1 \times \cdots \times A_n$ , where  $A_i$  are Lebesgue measurable in  $\mathbb{R}$ , then we get the product measure of the Lebesgue measure on  $\mathbb{R}$ . Of course we may also consider simpler rectangles. They all give the same measure, which we designate as the Lebesgue measure on  $\mathbb{R}^n$ .

**Proposition 11.4.1.** *The outer measures induced by the volume of the following collections are the same.*

1. Measurable rectangles.
2. Rectangles.
3. Open rectangles.
4. Open cubes.

*Proof.* We prove for  $n = 2$  only. The general case is similar.

Denote by  $\mu_1^*, \mu_2^*, \mu_3^*, \mu_4^*$  the outer measures generated by the four collections. Since the same volume on larger collection induces smaller outer measure, we have  $\mu_1^* \leq \mu_2^* \leq \mu_3^* \leq \mu_4^*$ . By Exercise 11.1.3, to show that  $\mu_4^* \leq \mu_3^* \leq \mu_1^*$ , it is sufficient to show that  $\mu_3^*(A \times B) \leq \mu(A)\mu(B)$  for Lebesgue measurable  $A, B \subset \mathbb{R}$ , and to show that  $\mu_4^*((a, b) \times (c, d)) \leq (b - a)(d - c)$ . Here and in the subsequent proof,  $\mu$  denotes the Lebesgue measure on  $\mathbb{R}$ .

Let  $A, B \subset \mathbb{R}$  be Lebesgue measurable. For any  $\epsilon > 0$ , there are open subsets  $U = \sqcup_i (a_i, b_i) \supset A$  and  $V = \sqcup_j (c_j, d_j) \supset B$ , such that  $\mu(U) < \mu(A) + \epsilon$  and  $\mu(V) < \mu(B) + \epsilon$ . Then  $A \times B \subset U \times V = \sqcup_{ij} (a_i, b_i) \times (c_j, d_j)$ , and we get

$$\begin{aligned} \mu_3^*(A \times B) &\leq \sum_{ij} (b_i - a_i)(d_j - c_j) = \left( \sum_i (b_i - a_i) \right) \left( \sum_j (d_j - c_j) \right) \\ &= \mu(U)\mu(V) \leq (\mu(A) + \epsilon)(\mu(B) + \epsilon). \end{aligned}$$

Since  $\epsilon$  is arbitrary, we get  $\mu_3^*(A \times B) \leq \mu(A)\mu(B)$ .

Strictly speaking, we still need to show that, if  $\mu(A) = +\infty$  and  $\mu(B) = 0$ , then  $\mu_3^*(A \times B) = 0$ . For any  $\epsilon > 0$ , we have open subsets  $U_i = \sqcup_j (a_{ij}, b_{ij}) \supset B$



satisfying  $\mu(U_i) < \frac{\epsilon}{i2^i}$ . We have

$$A \times B \subset \mathbb{R} \times B \subset \cup_i (-i, i) \times U_i = \cup_{ij} (-i, i) \times (a_{ij}, b_{ij}),$$

and

$$\mu_3^*(A \times B) \leq \sum_{ij} 2i(b_{ij} - a_{ij}) = \sum_i 2i\mu(U_i) < \sum_i 2i \frac{\epsilon}{i2^i} = 2\epsilon.$$

Since  $\epsilon$  is arbitrary, we get  $\mu_3^*(A \times B) = 0$ .

Next we prove  $\mu_4^*((a, b) \times (c, d)) \leq (b - a)(d - c)$ . For any  $\epsilon > 0$ , we have integers  $k, l \geq 0$ , such that

$$(k - 1)\epsilon \leq b - a < k\epsilon, \quad (l - 1)\epsilon \leq d - c < l\epsilon.$$

Then we can find open intervals  $(a_i, a_i + \epsilon)$ ,  $i = 1, 2, \dots, k$ , and open intervals  $(c_j, c_j + \epsilon)$ ,  $j = 1, 2, \dots, l$ , such that

$$(a, b) \subset \cup_{i=1}^k (a_i, a_i + \epsilon), \quad (c, d) \subset \cup_{j=1}^l (c_j, c_j + \epsilon).$$

Then  $(a, b) \times (c, d) \subset \cup_{ij} (a_i, a_i + \epsilon) \times (c_j, c_j + \epsilon)$ , and the right side is a union of open cubes of side length  $\epsilon$ . Therefore

$$\mu_4^*((a, b) \times (c, d)) \leq \sum_{ij} \epsilon^2 = (k\epsilon)(l\epsilon) < (b - a + \epsilon)(d - c + \epsilon).$$

Since  $\epsilon$  is arbitrary, we get  $\mu_4^*((a, b) \times (c, d)) \leq (b - a)(d - c)$ . □

**Exercise 11.25.** Prove that the Lebesgue measure is induced by the outer measure generated by the volume of the following collections of subsets.

1. Borel measurable rectangles: each side is Borel measurable in  $\mathbb{R}$ .
2. Rectangles.
3. Closed rectangles.
4. Rectangles of the form  $[a_1, b_1) \times \dots \times [a_n, b_n)$ .
5. Closed cubes.
6. Cubes of side length  $< 1$ .

**Exercise 11.26.** Explain that the outer measure generated by the volume of the collection of open cubes of side length 1 does not induce the Lebesgue measure.

**Exercise 11.27.** Let  $f$  be a continuous function on a closed bounded rectangle  $I \subset \mathbb{R}^{n-1}$ . Let  $A \subset \mathbb{R}^{n-1} \times \mathbb{R} = \mathbb{R}^n$  be its graph.

1. Prove that for any  $\epsilon > 0$ , there is a partition  $I = \cup I_i$  into finitely many closed rectangles, such that  $\mu_{n-1}(I) = \sum \mu_{n-1}(I_i)$  and  $\omega_{I_i}(f) \leq \epsilon$ .
2. Use the first part to construct  $A \subset \cup I_i \times [a_i, a_i + \epsilon]$ .
3. Prove that  $A$  has Lebesgue measure 0.

**Exercise 11.28.** Use Exercise 11.27 to prove that any submanifold of  $\mathbb{R}^n$  dimension  $< n$  has its  $n$ -dimensional Lebesgue measure equal to 0.

## Borel Set

**Definition 11.4.2.** The *Borel  $\sigma$ -algebra* is the smallest  $\sigma$ -algebra that contains all open subsets. The subsets in the Borel  $\sigma$ -algebra are *Borel sets*.

The reason for using open subsets instead of open rectangles is that the definition can be applied to more general topological spaces, not just Euclidean spaces.

**Example 11.4.1.** Since open cubes are exactly balls in the  $L^\infty$ -norm, by Proposition 6.4.5, a  $\sigma$ -algebra containing all open cubes must contain all open subsets. Therefore the smallest  $\sigma$ -algebra containing all open cubes is the Borel  $\sigma$ -algebra.

By taking  $L^2$ -norm instead of  $L^\infty$ -norm, the same idea shows that the Borel  $\sigma$ -algebra is also the smallest  $\sigma$ -algebra containing all open Euclidean balls.

**Exercise 11.29.** Prove that the Borel  $\sigma$ -algebra is the smallest  $\sigma$ -algebra containing any of the following collections.

1. Open rectangles.
2. Closed rectangles.
3. Rectangles of the form  $(a_1, b_1] \times \cdots \times (a_n, b_n]$ .
4. Open  $L^1$ -balls.
5. Closed cubes.
6. Closed subsets.
7. Open cubes of side length  $< 1$ .
8. Open cubes of side length 1.

**Exercise 11.30.** Prove that the Borel  $\sigma$ -algebra on  $\mathbb{R}^2$  is the smallest  $\sigma$ -algebra containing all open triangles.

**Exercise 11.31.** Prove that the Borel  $\sigma$ -algebra on  $\mathbb{R}^{m+n}$  is the product of the Borel  $\sigma$ -algebras on  $\mathbb{R}^m$  and  $\mathbb{R}^n$ .

As complements of open subsets, closed subsets are Borel sets. Then countable intersections of open subsets (called  $G_\delta$ -sets) and countable unions of closed subsets (called  $F_\sigma$ -sets) are also Borel sets. For example, countable subsets are  $F_\sigma$ -sets.

Furthermore, countable unions of  $G_\delta$ -sets (called  $G_{\delta\sigma}$ -sets) and countable intersections of  $F_\sigma$ -sets (called  $F_{\sigma\delta}$ -sets) are Borel sets.

Since a subset is open if and only if its complement is closed, a subset is  $G_\delta$  if and only if its complement is  $F_\sigma$ , and a subset is  $G_{\delta\sigma}$  if and only if its complement is  $F_{\sigma\delta}$ .

**Example 11.4.2.** The concept of semicontinuous functions in Example 10.2.3 can be easily extended to multivariable functions. In particular,  $f(\vec{x})$  is lower semicontinuous if and only if  $f^{-1}(a, +\infty)$  is open. Then for a sequence  $\epsilon_i > 0$  converging to 0, we have

$$f^{-1}[a, +\infty) = \bigcap_{i=1}^{\infty} f^{-1}(a - \epsilon_i, +\infty),$$

which is a  $G_\delta$ -set. This implies that  $f^{-1}(-\infty, a) = \mathbb{R} - f^{-1}[a, +\infty)$  is an  $F_\sigma$ -set.

**Example 11.4.3.** The set of continuous points of any function  $f$  on  $\mathbb{R}^n$  is a  $G_\delta$ -set.

Consider the oscillation of  $f$  on the ball  $B(\vec{x}, \delta)$

$$\omega_{B(\vec{x}, \delta)}(f) = \sup_{\substack{\|\vec{y} - \vec{x}\| < \delta \\ \|\vec{z} - \vec{x}\| < \delta}} |f(\vec{y}) - f(\vec{z})| = \sup_{B(\vec{x}, \delta)} f - \inf_{B(\vec{x}, \delta)} f.$$

If  $\omega_{B(\vec{x}, \delta)}(f) < \epsilon$  and  $\vec{y} \in B(\vec{x}, \delta)$ , then  $B(\vec{y}, \delta') \subset B(\vec{x}, \delta)$  for  $\delta' = \delta - \|\vec{y} - \vec{x}\| > 0$ , so that  $\omega_{B(\vec{y}, \delta')}(f) \leq \omega_{B(\vec{x}, \delta)}(f) < \epsilon$ . This shows that for any  $\epsilon > 0$ , the subset

$$U_\epsilon = \{\vec{x} : \omega_{B(\vec{x}, \delta)}(f) < \epsilon \text{ for some } \delta > 0\}$$

is open. Then for a sequence  $\epsilon_i > 0$  converging to 0, the  $G_\delta$ -set

$$\cap_{i=1}^\infty U_{\epsilon_i} = \{\vec{x} : \omega_{\vec{x}}(f) = \lim_{\delta \rightarrow 0} \omega_{B(\vec{x}, \delta)}(f) = 0\},$$

is exactly the set of continuous points of the function. See Exercise 4.85.

**Example 11.4.4.** The set of differentiable points of any continuous function  $f$  on  $\mathbb{R}$  is an  $F_{\sigma\delta}$ -set.

For  $t \neq 0$ , the function  $D_t(x) = \frac{f(x+t) - f(x)}{t}$  is continuous. Let  $\epsilon_i > 0$  converge to 0, then the differentiability at  $x$  means the “upper and lower” derivatives

$$\begin{aligned} \overline{\lim}_{t \rightarrow 0} D_t(x) &= \inf_i \sup_{0 < |t| < \epsilon_i} D_t(x), & g_i(x) &= \sup_{0 < |t| < \epsilon_i} D_t(x), \\ \underline{\lim}_{t \rightarrow 0} D_t(x) &= \sup_i \inf_{0 < |t| < \epsilon_i} D_t(x), & h_i(x) &= \inf_{0 < |t| < \epsilon_i} D_t(x), \end{aligned}$$

are equal. Therefore the set of non-differentiable points is

$$\begin{aligned} \{x : \inf_i g_i(x) > \sup_i h_i(x)\} &= \cup_{r \in \mathbb{Q}} \{x : \inf_i g_i(x) > r > \sup_i h_i(x)\} \\ &= \cup_{r \in \mathbb{Q}} (\{x : \inf_i g_i(x) > r\} \cap \{x : \sup_i h_i(x) < r\}). \end{aligned}$$

As the supremum of continuous functions,  $g_i$  is lower semicontinuous. Therefore  $\{x : g_i(x) > a\}$  is open, and

$$\{x : \inf_i g_i(x) > r\} = \cup_j (\cap_i \{x : g_i(x) > r + \epsilon_j\})$$

is a  $G_{\delta\sigma}$ -set. Similarly,  $\{x : \sup_i h_i(x) < r\}$  is also a  $G_{\delta\sigma}$ -set. The intersection of the two  $G_{\delta\sigma}$ -sets is still a  $G_{\delta\sigma}$ -set. As a countable union of  $G_{\delta\sigma}$ -sets, the set of non-differentiable points is still a  $G_{\delta\sigma}$ -set. The set of non-differentiable points is then a complement of the  $G_{\delta\sigma}$ -set, and is therefore an  $F_{\sigma\delta}$ -set.

**Exercise 11.32.** Use Exercise 6.80 to prove that any closed subset is a  $G_\delta$ -set. This implies that any open subset is an  $F_\sigma$ -set.

**Exercise 11.33.** Suppose  $f$  is lower semicontinuous and  $g$  is upper semicontinuous. Determine the type of subsets.

- |                           |                              |
|---------------------------|------------------------------|
| 1. $f^{-1}(-\infty, a)$ . | 4. $\{x: f(x) > g(x)\}$ .    |
| 2. $f^{-1}(a, b)$ .       | 5. $\{x: g(x) > f(x)\}$ .    |
| 3. $g^{-1}(a, b)$ .       | 6. $\{x: f(x) \geq g(x)\}$ . |

**Exercise 11.34.** Suppose  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a continuous map. Prove that for any Borel set  $A \subset \mathbb{R}^m$ , the preimage  $F^{-1}(A)$  is a Borel set in  $\mathbb{R}^n$ .

**Exercise 11.35.** Prove that a monotone functions on a Borel set  $A \subset \mathbb{R}$  is Borel measurable. Moreover, monotone functions on a Lebesgue measurable subset are Lebesgue measurable.

**Exercise 11.36.** Prove that for countably many functions, the set of points where all the functions are continuous is a Borel set. What about the points where at least one function is continuous? What about the points where infinitely many functions are continuous?

Borel sets are Lebesgue measurable. The following result shows that Borel sets can be used to approximate Lebesgue measurable subsets, and the Lebesgue measure space is the completion of the Borel measure space.

**Proposition 11.4.3.** *The following are equivalent to the Lebesgue measurability of a subset  $A \subset \mathbb{R}^n$ .*

1. For any  $\epsilon > 0$ , there is an open subset  $U$  and a closed subset  $C$ , such that  $C \subset A \subset U$  and  $\mu(U - C) < \epsilon$ .
2. There is a  $G_\delta$ -set  $D$  and an  $F_\sigma$ -set  $S$ , such that  $S \subset A \subset D$ , and  $\mu(D - S) = 0$ .

Moreover, in case  $\mu^*(A)$  is finite, we may take  $C$  in the first statement to be compact.

*Proof.* We largely follow the proof of Proposition 11.2.1.

Suppose  $A$  is Lebesgue measurable. For any  $\epsilon > 0$ , there is  $A \subset U = \cup I_i$ , such that  $\sum \mu(I_i) < \mu(A) + \epsilon$ . Here by Proposition 11.4.1, we may choose  $I_i$  to be open rectangles. Then  $U$  is open and, when  $\mu(A) < +\infty$ , we have

$$\mu(U - A) = \mu(U) - \mu(A) \leq \sum \mu(I_i) - \mu(A) < \epsilon.$$

For general  $A$ , we have  $A = \cup A_j$  with  $\mu(A_j) < +\infty$  (take  $A_j = A \cap [-j, j]^n$ , for example). Then we have open  $U_j \supset A_j$ , such that  $\mu(U_j - A_j) < \frac{\epsilon}{2^j}$ . This implies that  $U = \cup U_j$  is an open subset containing  $A$ , such that

$$\mu(U - A) = \mu(\cup (U_j - A)) \leq \mu(\cup (U_j - A_j)) \leq \sum \mu(U_j - A_j) < \sum \frac{\epsilon}{2^j} = \epsilon.$$

Applying the conclusion to the Lebesgue measurable subset  $\mathbb{R}^n - A$ , there is an open  $V \supset \mathbb{R}^n - A$ , such that  $\mu(V - (\mathbb{R}^n - A)) < \epsilon$ . Then  $C = \mathbb{R}^n - V$  is a closed subset contained in  $A$ , and  $V - (\mathbb{R}^n - A) = A - C$ , so that  $\mu(A - C) < \epsilon$ . Thus we found open  $U$  and closed  $C$ , such that  $C \subset A \subset U$  and  $\mu(U - C) \leq \mu(U - A) + \mu(A - C) < 2\epsilon$ .

In case  $\mu(A) < +\infty$ , we have  $\lim_{r \rightarrow +\infty} \mu(A - C \cap [-r, r]^n) = \mu(A - C)$ . By replacing  $C$  with  $C \cap [-r, r]^n$  for sufficiently big  $r$ , we still have  $\mu(A - C) < \epsilon$ , but with a compact  $C$ .

Now assume the first statement. We have sequences of open subsets  $U_j$  and closed subsets  $C_j$ , such that  $C_j \subset A \subset U_j$  and  $\lim \mu(U_j - C_j) = 0$ . Then the  $G_\delta$ -set  $D = \cap U_j$  and  $F_\sigma$ -set  $S = \cup C_j$  satisfy  $S \subset A \subset D$  and  $C_j \subset S \subset D \subset U_j$  for all  $j$ . This implies  $\mu(D - S) \leq \mu(U_j - C_j)$ . Then  $\lim \mu(U_j - C_j) = 0$  implies  $\mu(D - S) = 0$ .

Finally assume the second statement. Then by the completeness of the Lebesgue measure,  $A - S \subset D - S$  and  $\mu(D - S) = 0$  imply that  $A - S$  is Lebesgue measurable. Therefore  $A = S \cup (A - S)$  is Lebesgue measurable.  $\square$

## Translation Invariant Measure

Fix a vector  $\vec{a} \in \mathbb{R}^n$ . For any  $A \subset \mathbb{R}^n$ , the subset

$$\vec{a} + A = \{\vec{a} + \vec{x} : \vec{x} \in A\}$$

is the *translation* of  $A$  by vector  $\vec{a}$ . The translation is actually defined in any vector space.

**Theorem 11.4.4.** *Let  $\nu$  be a translation invariant measure on the Lebesgue  $\sigma$ -algebra of a finite dimensional vector space, such that  $\nu(A)$  is finite for any bounded subset  $A$ . Then  $\nu$  is a constant multiple of the Lebesgue measure.*

The theorem is stated for finite dimensional vector space because we want to emphasize the independent of the result from the coordinate system. More generally, the translation can be extended to group actions, and we may ask similar question of measures invariant under group actions. For the case of groups acting on itself (and bounded replaced by compact), the extension of the theorem gives the *Haar measure* that is unique up to a constant scalar.

*Proof.* Because any finite dimensional vector space is isomorphic to some  $\mathbb{R}^n$ , we only need to prove the theorem for  $\mathbb{R}^n$ . To simplify the presentation, we will prove for  $n = 2$  only. The general case is similar.

Let  $\mu$  be the Lebesgue measure, defined as the product of the Lebesgue measure on  $\mathbb{R}$ . Denote the square  $I_\epsilon = (0, \epsilon] \times (0, \epsilon]$ . By multiplying a constant number if necessary, we assume  $\nu(I_1) = 1$ . We will show that  $\nu = \mu$ .

For a natural number  $n$ , the unit square  $I_1$  is the disjoint union of  $n^2$  copies of translations of the small square  $I_{\frac{1}{n}}$ . By translation invariance and additivity, we have  $1 = \nu(I_1) = n^2 \nu(I_{\frac{1}{n}})$ . Therefore  $\nu(I_{\frac{1}{n}}) = \frac{1}{n^2}$ .

For any rational numbers  $a < b$  and  $c < d$ , we have  $b - a = \frac{k}{n}$ ,  $d - c = \frac{l}{n}$  for some natural numbers  $k, l, n$ , and the rectangle  $(a, b] \times (c, d]$  is the disjoint union of  $kl$  copies of translations of  $I_{\frac{1}{n}}$ . By translation invariance and additivity, we have  $\nu((a, b] \times (c, d]) = kl \nu(I_{\frac{1}{n}}) = \frac{kl}{n^2} = (b - a)(d - c)$ .

For any rectangle  $\langle a, b \rangle \times \langle c, d \rangle$ , we consider  $(a_1, b_1] \subset \langle a, b \rangle \subset (a_2, b_2]$  and  $(c_1, d_1] \subset \langle c, d \rangle \subset (c_2, d_2]$  for rational  $a_i, b_i, c_i, d_i$ . We have

$$\nu((a_1, b_1] \times (c_1, d_1]) \leq \nu(\langle a, b \rangle \times \langle c, d \rangle) \leq \nu((a_2, b_2] \times (c_2, d_2]).$$

When  $a_i \rightarrow a$ ,  $b_i \rightarrow b$ ,  $c_i \rightarrow c$ ,  $d_i \rightarrow d$ , the left side  $\nu((a_1, b_1] \times (c_1, d_1]) = (b_1 - a_1)(d_1 - c_1)$  converges to  $(b - a)(d - c)$ , and the right side also converges to  $(b - a)(d - c)$ . Therefore  $\nu(\langle a, b \rangle \times \langle c, d \rangle) = (b - a)(d - c) = \mu(\langle a, b \rangle \times \langle c, d \rangle)$ .

Consider a Lebesgue measurable rectangle  $A \times B$ . Since  $A$  and  $B$  are Lebesgue measurable in  $\mathbb{R}$ , for any  $\epsilon > 0$ , we have  $A \subset U = \sqcup (a_i, b_i)$  and  $B \subset V = \sqcup (c_j, d_j)$ , such that  $(\mu_1$  is the Lebesgue measure on  $\mathbb{R})$

$$\mu_1(U) = \sum (b_i - a_i) < \mu_1(A) + \epsilon, \quad \mu_1(V) = \sum (d_j - c_j) < \mu_1(B) + \epsilon.$$

Then by  $A \times B \subset U \times V = \sqcup (a_i, b_i) \times (c_j, d_j)$ , we have

$$\begin{aligned} \nu(A \times B) &\leq \sum \nu((a_i, b_i) \times (c_j, d_j)) = \sum \mu((a_i, b_i) \times (c_j, d_j)) \\ &= \mu(U \times V) = \mu_1(U)\mu_1(V) \leq (\mu_1(A) + \epsilon)(\mu_1(B) + \epsilon). \end{aligned}$$

Since  $\epsilon$  is arbitrary, we get  $\nu(A \times B) \leq \mu_1(A)\mu_1(B) = \mu(A \times B)$ . Although we can further prove the equality, the equality is not needed for the moment.

Now consider any Lebesgue measurable subset  $A$  of  $\mathbb{R}^2$ . For any  $\epsilon > 0$ , there are countably many Lebesgue measurable rectangles  $I_i$ , such that  $A \subset \cup I_i$  and  $\sum \mu(I_i) \leq \mu(A) + \epsilon$ . Then

$$\nu(A) \leq \sum \nu(I_i) \leq \sum \mu(I_i) \leq \mu(A) + \epsilon.$$

Here we just proved the second inequality. Since  $\epsilon$  is arbitrary, we get  $\nu(A) \leq \mu(A)$ . For bounded  $A$ , we have  $A \subset I$  for some rectangle  $I$ . Then  $\nu(I) = \mu(I)$ ,  $\nu(I - A) \leq \mu(I - A)$ , so that

$$\nu(A) = \nu(I) - \nu(I - A) \geq \mu(I) - \mu(I - A) = \mu(A).$$

Thus  $\nu(A) = \mu(A)$  for bounded and Lebesgue measurable  $A$ . The general case can be obtained by applying the monotone limit property in Proposition 9.4.4 to  $A \cap [-n, n] \times [-n, n]$  for increasing  $n$ .  $\square$

## Lebesgue Measure under Linear Transform

**Proposition 11.4.5.** *A linear transform  $L: \mathbb{R}^n \rightarrow \mathbb{R}^n$  takes Lebesgue measurable subsets to Lebesgue measurable subsets, and*

$$\mu(L(A)) = |\det(L)|\mu(A).$$

*Proof.* By Proposition 6.3.3, a continuous map takes compact subsets to compact subsets. By Exercise 6.63, any closed subset is union of countably many compact subsets. Therefore the image of a closed subset is the union of countably many

compact subsets. By Proposition 6.3.6, compact implies closed. Therefore the image is also an  $F_\sigma$ -set. This further implies that the image of any  $F_\sigma$ -set is an  $F_\sigma$ -set.

If  $A$  is Lebesgue measurable, then by Proposition 11.4.3, there is an  $F_\sigma$ -set  $S \subset A$  satisfying  $\mu(A - S) = 0$ . Then  $L(A) = L(S) \cup L(A - S)$ . We know  $L(S)$  is an  $F_\sigma$ -set and is therefore Lebesgue measurable. If we can show that the outer measure of  $L(A - S)$  is zero, then  $L(A - S)$  is also Lebesgue measurable, so that  $L(A)$  is Lebesgue measurable.

So we need to show that  $\mu(A) = 0$  implies  $\mu(L(A)) = 0$ . For any  $\epsilon > 0$ , by Theorem 11.4.1, there are countably many cubes  $I_i$ , such that  $A \subset \cup I_i$  and  $\sum \mu(I_i) < \epsilon$ . The cubes are  $L^\infty$ -balls  $I_i = B_{L^\infty}(\vec{a}_i, \epsilon_i)$ . On the other hand, we have

$$\|L(\vec{x}) - L(\vec{y})\|_\infty \leq \|L\| \|\vec{x} - \vec{y}\|_\infty,$$

where the norm  $\|L\|$  is taken with respect to the  $L^\infty$ -norm on  $\mathbb{R}^n$ . The inequality implies that

$$L(I_i) = L(B_{L^\infty}(\vec{a}_i, \epsilon_i)) \subset B_{L^\infty}(L(\vec{a}_i), \|L\| \epsilon_i).$$

Therefore

$$\mu(L(I_i)) \leq \mu(B_{L^\infty}(L(\vec{a}_i), \|L\| \epsilon_i)) = (2\|L\| \epsilon_i)^n = \|L\|^n (2\epsilon_i)^n = \|L\|^n \mu(I_i).$$

By  $L(A) \subset \cup L(I_i)$ , we get

$$\mu^*(L(A)) \leq \sum \mu^*(L(I_i)) \leq \|L\|^n \sum \mu(I_i) < \|L\|^n \epsilon.$$

Since  $\epsilon$  is arbitrary, we get  $\mu^*(L(A)) = 0$ .

After proving that  $L$  preserves Lebesgue measurable subsets, the composition  $\mu_L(A) = \mu(L(A))$  becomes a function on the  $\sigma$ -algebra of Lebesgue measurable subsets. If  $L$  is invertible, then  $L$  preserves disjoint unions, so that  $\mu_L$  is a measure. Moreover,  $\mu_L$  is translation invariant

$$\mu_L(\vec{a} + A) = \mu(L(\vec{a}) + L(A)) = \mu(L(A)) = \mu_L(A).$$

Therefore by Theorem 11.4.4, we get

$$\mu(L(A)) = c\mu(A), \quad c = \mu(L((0, 1]^n)).$$

Applying the equality to  $A = (0, 1]^n$  and using Theorem 7.4.2, we get  $c = |\det(L)|$ .  $\square$

## Non-Lebesgue Measurable Subsets

We may use the translation invariance to construct subsets that are not Lebesgue measurable. We also note that by Exercise 11.42, not all Lebesgue measurable subsets are Borel sets.

**Theorem 11.4.6.** *Every subset with positive Lebesgue outer measure contains a subset that is not Lebesgue measurable.*

*Proof.* Let  $\mathbb{Q}^n$  be the collection of all rational vectors in  $\mathbb{R}^n$ . The first step is to construct a subset  $X \subset (0, 1)^n$ , such that the sum map

$$+: \mathbb{Q}^n \times X \rightarrow \mathbb{R}^n$$

is a one-to-one correspondence.

The translation  $\mathbb{Q}^n + \vec{x}$  is the collection of vectors that differ from  $\vec{x}$  by rational vectors. For any  $\vec{x}$  and  $\vec{y}$ , only two mutually exclusive cases may happen:

- If  $\vec{x} - \vec{y} \in \mathbb{Q}^n$ , then  $\mathbb{Q}^n + \vec{x} = \mathbb{Q}^n + \vec{y}$ .
- If  $\vec{x} - \vec{y} \notin \mathbb{Q}^n$ , then  $\mathbb{Q}^n + \vec{x}$  and  $\mathbb{Q}^n + \vec{y}$  are disjoint.

Therefore the subsets of the form  $\mathbb{Q}^n + \vec{x}$  form a disjoint union decomposition of  $\mathbb{R}^n$ . Choose one vector in each subset  $\mathbb{Q}^n + \vec{x}$  in the disjoint union decomposition. Since  $(\mathbb{Q}^n + \vec{x}) \cap (0, 1)^n$  is not empty, we may further assume that the vector lies in  $(0, 1)^n$ . All the vectors we choose form a subset  $X \subset (0, 1)^n$ , and the disjoint union decomposition means

$$\mathbb{R}^n = \sqcup_{\vec{x} \in X} (\mathbb{Q}^n + \vec{x}) = \sqcup_{\vec{r} \in \mathbb{Q}^n} (\vec{r} + X).$$

This is equivalent to that the sum map  $\mathbb{Q}^n \times X \rightarrow \mathbb{R}^n$  is a one-to-one correspondence.

Next we prove that  $X$  is not Lebesgue measurable. The proof is not necessary for the general conclusion of the theorem, but provides the key idea of the argument. Choose distinct rational vectors  $\vec{r}_i \in \mathbb{Q}^n \cap (0, 1)^n$ ,  $i = 1, \dots, N$ . If  $X$  is Lebesgue measurable, then the translations  $\vec{r}_i + X$  are also Lebesgue measurable, with  $\mu(\vec{r}_i + X) = \mu(X)$ . Moreover,  $\vec{r}_i + X$  are disjoint and contained in  $(0, 1)^n + (0, 1)^n \subset (0, 2)^n$ . Therefore

$$\begin{aligned} N\mu(X) &= \mu(\vec{r}_1 + X) + \dots + \mu(\vec{r}_N + X) \\ &= \mu((\vec{r}_1 + X) \sqcup \dots \sqcup (\vec{r}_N + X)) \\ &\leq \mu((0, 2)^n) = 2^n. \end{aligned}$$

Since  $N$  can be arbitrarily big, we get  $\mu(X) = 0$ . Then by the countable union  $\mathbb{R}^n = \sqcup_{\vec{r} \in \mathbb{Q}^n} (\vec{r} + X)$  and  $\mu(\vec{r} + X) = \mu(X)$ , we get  $\mu(\mathbb{R}^n) = 0$ , a contradiction.

In general, let  $A$  be a subset with  $\mu^*(A) > 0$ . Then

$$A = A \cap (\sqcup_{\vec{r} \in \mathbb{Q}^n} (\vec{r} + X)) = \sqcup_{\vec{r} \in \mathbb{Q}^n} (A \cap (\vec{r} + X)).$$

Suppose for a specific  $\vec{r}$ , the subset  $B = A \cap (\vec{r} + X) \subset \vec{r} + X \subset \vec{r} + (0, 1)^n$  is measurable. Again choose distinct  $\vec{r}_i \in \mathbb{Q}^n \cap (0, 1)^n$ . Then we get disjoint measurable subsets  $\vec{r}_i + B \subset (0, 1)^n + \vec{r} + (0, 1)^n \subset \vec{r} + (0, 2)^n$ . Therefore

$$\begin{aligned} N\mu(B) &= \mu(\vec{r}_1 + B) + \dots + \mu(\vec{r}_N + B) \\ &= \mu((\vec{r}_1 + B) \sqcup \dots \sqcup (\vec{r}_N + B)) \\ &\leq \mu(\vec{r} + (0, 2)^n) = 2^n. \end{aligned}$$

Since  $N$  can be arbitrary, we get  $\mu(B) = 0$ .



If  $A \cap (\vec{r} + X)$  is measurable for all  $\vec{r} \in Q$ , then by what we just proved, each term in the countable union  $A = \sqcup_{\vec{r} \in Q} (A \cap (\vec{r} + X))$  is measurable and has zero measure. This implies that  $\mu^*(A) = \mu(A) = 0$ . The contradiction to the assumption  $\mu^*(A) > 0$  shows that at least one subset  $A \cap (\vec{r} + X)$  of  $A$  is not measurable.  $\square$

**Example 11.4.5 (Cantor Function).** The cantor set in Example 9.2.4 was obtained by deleting the middle third intervals successively. The *Cantor function*  $\kappa$  is an increasing function obtained by elevating the middle third intervals successively into half positions. Specifically, we have

$$\kappa = \frac{1}{2} = 0.1_{[2]} \text{ on } \left(\frac{1}{3}, \frac{2}{3}\right) = (0.1_{[3]}, 0.2_{[3]}).$$

Then we elevate the next middle third intervals. The left interval  $\left(\frac{1}{9}, \frac{2}{9}\right)$  is elevated to  $\frac{1}{4}$ , midway between 0 and  $\frac{1}{2}$ . The right interval  $\left(\frac{7}{9}, \frac{8}{9}\right)$  is elevated to  $\frac{3}{4}$ , midway between  $\frac{1}{2}$  and 1. The result is

$$\kappa = 0.01_{[2]} \text{ on } (0.01_{[3]}, 0.02_{[3]}), \quad \kappa = 0.11_{[2]} \text{ on } (0.21_{[3]}, 0.22_{[3]}).$$

Keep going, we have

$$\kappa = 0.b_1b_2 \cdots b_n 1_{[2]} \text{ on } (0.a_1a_2 \cdots a_n 1_{[3]}, 0.a_1a_2 \cdots a_n 2_{[3]}), \quad a_i = 0 \text{ or } 2, \quad b_i = \frac{a_i}{2}.$$

This gives the value of  $\kappa$  on the complement of the Cantor set  $K$ .

The value of  $\kappa$  on  $K$  can be determined by the continuity. By Theorem 2.4.1, such continuity must be uniform. Moreover, the continuity implies the values at the ends of the deleted open intervals

$$\kappa(0.a_1a_2 \cdots a_n 1_{[3]}) = \kappa(0.a_1a_2 \cdots a_n 2_{[3]}) = 0.b_1b_2 \cdots b_n 1_{[2]}.$$

Now consider  $x = 0.a_1a_2 \cdots a_n \cdots_{[3]} \in K$ , which means  $a_i = 0$  or  $2$ . By  $|x - 0.a_1a_2 \cdots a_n 1_{[3]}| \leq \frac{1}{3^n}$  and the uniform continuity, for any  $\epsilon > 0$ , there is  $N$ , such that

$$n > N \implies |\kappa(x) - \kappa(0.a_1a_2 \cdots a_n 1_{[3]})| \leq \epsilon.$$

If there are infinitely many  $a_n = 2$ , then we can always find  $a_n = 2$  with  $n > N$ , and we get (note  $b_n = 1$ )

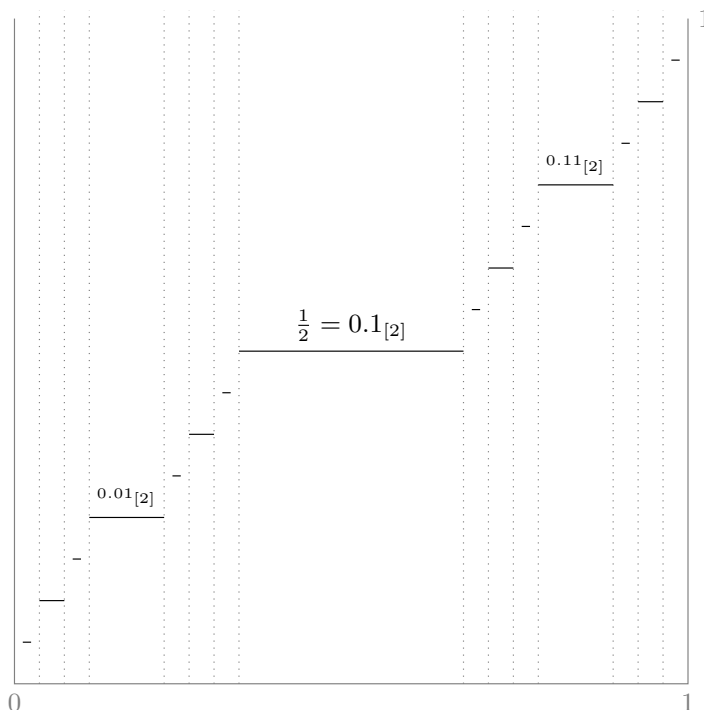
$$\begin{aligned} |\kappa(x) - 0.b_1b_2 \cdots b_n \cdots_{[2]}| &\leq |\kappa(x) - 0.b_1b_2 \cdots b_{n-1}b_n 1_{[2]}| + \frac{1}{2^n} \\ &= |\kappa(x) - \kappa(0.a_1a_2 \cdots a_{n-1}a_n 1_{[3]})| + \frac{1}{2^n} < \epsilon + \frac{1}{2^N}. \end{aligned}$$

Since  $\epsilon$  is arbitrary and  $N$  can be as large as we wish, this implies

$$\kappa(0.a_1a_2 \cdots a_n \cdots_{[3]}) = 0.b_1b_2 \cdots b_n \cdots_{[2]} \text{ on } K.$$

If there are only finitely many 2 among  $a_n$ , then  $x = 0.a_1a_2 \cdots a_n 2_{[3]}$ , and the formula remains valid.

Although we determined the value of  $\kappa$  by using continuity, we still need to verify that the function constructed above is indeed continuous. This is left as Exercise 11.39.



**Figure 11.4.1.** *Cantor function*

We also note that the value of  $\kappa$  on  $K$  can also be determined by the increasing property. See Exercise 11.40. Therefore  $\kappa$  is an increasing and continuous function.

Let  $A \subset [0, 1]$  be a Lebesgue non-measurable subset. Then there is (almost unique)  $B \subset K$ , such that  $\kappa(B) = A$ . Since  $\mu(K) = 0$ , the subset  $B$  is always Lebesgue measurable. Thus we see that the image of a Lebesgue measurable subset under a continuous map is not necessarily Lebesgue measurable.

Even if the continuous map is invertible, it may still send a Lebesgue measurable subset to a Lebesgue non-measurable subset. See Exercise 11.41.

**Exercise 11.37.** Suppose  $\mu$  is a measure on  $Z = Y \times X$ , such that  $\mu(y \times X) = \mu(y' \times X)$  for any  $y, y' \in Y$ . If  $0 < \mu(Z) < +\infty$  and  $Y$  is countable, prove that every measurable subset of  $X$  must have measure 0, and  $X$  cannot be measurable.

**Exercise 11.38.** Let  $A$  be the set of end points of the deleted middle third intervals in the construction of the Cantor set  $K$ . Show that  $\kappa(A) \cap \kappa(K - A) = \emptyset$ ,  $\kappa|_A$  is two-to-one, and  $\kappa|_{K-A}$  is one-to-one.

**Exercise 11.39.** Show that if the first  $n$  digits of the base 3 expressions of  $x, y \in [0, 1]$  are the same, then the first  $n$  digits of the base 2 expressions of  $\kappa(x)$  and  $\kappa(y)$  are the same. Then use this to prove that the Cantor function is continuous.

**Exercise 11.40.** Prove that the Cantor function is also defined by its values on  $[0, 1] - K$

and the increasing property. Moreover,  $\kappa|_K$  is strictly increasing.

**Exercise 11.41.** Consider  $\phi(x) = \kappa(x) + x$ .

1. Show that  $\phi$  is a continuous invertible map from  $[0, 1]$  to  $[0, 2]$ .
2. Show that  $\phi(K)$  and  $\phi([0, 1] - K)$  are measurable subsets of  $[0, 2]$  of measure 1.
3. Show that there is a Lebesgue measurable  $A \subset [0, 1]$ , such that  $\phi(A)$  is not Lebesgue measurable.

The last statement means that the preimage of  $A$  under the continuous map  $\phi^{-1}: [0, 2] \rightarrow [0, 1]$  is not measurable.

**Exercise 11.42.** Use Exercise 11.34 and Exercise 11.41 to show that Lebesgue measurable subsets there are not Borel measurable.

**Exercise 11.43.** Use Exercise 11.41 to construct a Lebesgue measurable function  $g$  and a continuous function  $h$ , such that the composition  $g \circ h$  is not Lebesgue measurable. The counterexample should be compared with the first property in Proposition 10.2.2.

## 11.5 Riemann Integration on $\mathbb{R}^n$

The Lebesgue integral is much more flexible than the Riemann integral, and have much better convergence property. It removes some artificial difficulties inherent in the Riemann integral and is the more rational integration theory. Given we have already established the Lebesgue integral, there is no more need for the Riemann integral. Still, we outline the theory of Riemann integral on the Euclidean space to give the reader a more comprehensive picture of the integration theories.

### Jordan Measure

It is easy to imagine how to extend the Riemann integration to multivariable functions on rectangles. However, integration on rectangles alone is too restrictive, a theory of the volume of nice subsets needs to be established. It was Jordan and Peano who established such a theory along the line of Riemann integral on an interval, before Lebesgue established his measure theory. Jordan and Peano's idea is to approximate a subset from inside and outside by *finite* unions of rectangles. Of course, Lebesgue's revolution is to consider approximation from outside by *countable* unions of rectangles. The allowance of countable union and the countable additivity is the key reason Lebesgue's theory is much more flexible.

Let  $A \subset \mathbb{R}^n$  be a bounded subset. Then  $A$  is contained in a big rectangle. By taking a partition of each coordinate interval of the rectangle and then taking the product of the intervals in these partitions, we get a (rectangular) partition  $P$  of the big rectangle. The size of the partition can be measured by

$$\|P\| = \max\{d(I) : I \in P\},$$

where for closed rectangle  $I = [a_1, b_1] \times \cdots \times [a_n, b_n]$ ,

$$d(I) = \sup\{\|\vec{x} - \vec{y}\|_\infty : \vec{x}, \vec{y} \in I\} = \max\{b_1 - a_1, b_2 - a_2, \dots, b_n - a_n\}.$$

The partition  $P$  gives us finite unions of rectangles that approximates  $A$  from inside and from outside

$$A_P^- = \cup\{I: I \in P, I \subset A\} \subset A \subset A_P^+ = \cup\{I: I \in P, I \cap A \neq \emptyset\} \quad (11.5.1)$$

Then the volume of  $A$  should lie between the following inner and outer approximations

$$\mu(A_P^-) = \sum_{I \in P, I \subset A} \mu(I), \quad \mu(A_P^+) = \sum_{I \in P, I \cap A \neq \emptyset} \mu(I), \quad (11.5.2)$$

where  $\mu(I)$  is the usual volume of rectangles.

**Definition 11.5.1.** The *inner volume* and the *outer volume* of  $A \subset \mathbb{R}^n$  are

$$\mu^-(A) = \sup_P \mu(A_P^-), \quad \mu^+(A) = \inf_P \mu(A_P^+).$$

A subset  $A \subset \mathbb{R}^n$  is *Jordan measurable* if  $\mu^+(A) = \mu^-(A)$ , and the common value is *Jordan measure* (or the *volume*) of  $A$ .

The inner and outer volumes are monotone

$$A \subset B \implies \mu^-(A) \leq \mu^-(B), \quad \mu^+(A) \leq \mu^+(B).$$

The outer volume is *subadditive* (but not countably subadditive)

$$\mu^+(A_1 \cup A_2 \cup \cdots \cup A_k) \leq \mu^+(A_1) + \mu^+(A_2) + \cdots + \mu^+(A_k). \quad (11.5.3)$$

Consequently, the Jordan measure is monotone and subadditive.

We have  $\mu(A_P^+) \geq \mu(A_Q^+) \geq \mu(A_Q^-) \geq \mu(A_P^-)$  for a refinement  $Q$  of  $P$ . Since any two partitions have a common refinement, we have  $\mu^+(A) \geq \mu^-(A)$ . The common refinement also tells us that  $A$  is Jordan measurable if and only if for any  $\epsilon > 0$ , there is a partition  $P$ , such that  $\mu(A_P^+) - \mu(A_P^-) < \epsilon$ .

**Example 11.5.1.** Consider  $A = (a, b) \subset \mathbb{R}$ . For any  $\epsilon > 0$ , take  $P$  to be  $a < a + \epsilon < b - \epsilon < b$ . Then  $\mu(A_P^-) = \mu([a + \epsilon, b - \epsilon]) = b - a - 2\epsilon$  and  $\mu(A_P^+) = \mu([a, b]) = b - a$ . Therefore  $(a, b)$  has Jordan measure  $b - a$ .

**Example 11.5.2.** Consider  $B = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\}$ . For any natural number  $N$ , take  $P$  to be the partition of  $[0, 1]$  with partition points  $\frac{i}{N^2}$ ,  $0 \leq i \leq N^2$ . Then  $B_P^+$  consists of  $\left[0, \frac{1}{N}\right]$  (the union of first  $N$  intervals in  $P$ ) and at most  $N - 1$  intervals of length  $\frac{1}{N^2}$  (to cover those  $\frac{1}{n}$  not inside  $\left[0, \frac{1}{N}\right]$ ). Therefore  $\mu(B_P^+) \leq \frac{1}{N} + (N - 1)\frac{1}{N^2} < \frac{2}{N}$ , and  $B$  has Jordan measure 0.

**Example 11.5.3.** Let  $A$  be the set of rational numbers in  $[0, 1]$ . Then any interval  $[a, b]$  with  $a < b$  contains points inside and outside  $A$ . Therefore for any partition  $P$  of  $[0, 1]$ , we have

$$A_P^+ = [0, 1], \quad A_P^- = \emptyset, \quad \mu(A_P^+) = 1, \quad \mu(A_P^-) = 0.$$

We conclude that  $A$  is not Jordan measure, despite being Lebesgue measurable.

**Example 11.5.4.** Let  $f(x) \geq 0$  be a bounded function on  $[a, b]$ . The subset

$$G(f) = \{(x, y) : a \leq x \leq b, 0 \leq y \leq f(x)\} \subset \mathbb{R}^2$$

is the subset under the graph of the function. For any partition  $P$  of  $[a, b]$  and any partition  $Q$  of the  $y$ -axis, we have

$$\begin{aligned} G(f)_{P \times Q}^- &= \cup [x_{i-1}, x_i] \times [0, y_l], & y_l &= \max\{y_j : y_j \leq f \text{ on } [x_{i-1}, x_i]\}, \\ G(f)_{P \times Q}^+ &= \cup [x_{i-1}, x_i] \times [0, y_u], & y_u &= \min\{y_j : y_j > f \text{ on } [x_{i-1}, x_i]\}. \end{aligned}$$

The definition of  $y_l$  tells us

$$y_l \leq \inf_{[x_{i-1}, x_i]} f(x) < y_{l+1} \leq y_l + \|Q\|.$$

Therefore

$$\mu(G(f)_{P \times Q}^-) = \sum y_l \Delta x_i \leq \sum \inf_{[x_{i-1}, x_i]} f(x) \Delta x_i \leq \mu(G(f)_{P \times Q}^-) + \|Q\|(b-a).$$

Similarly, we have

$$\mu(G(f)_{P \times Q}^+) \geq \sum \sup_{[x_{i-1}, x_i]} f(x) \Delta x_i \geq \mu(G(f)_{P \times Q}^+) - \|Q\|(b-a).$$

Thus

$$0 \leq \mu(G(f)_{P \times Q}^+) - \mu(G(f)_{P \times Q}^-) - \sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i \leq 2\|Q\|(b-a).$$

Since  $\|Q\|$  can be arbitrarily small, we see that  $\mu(G(f)_{P \times Q}^+) - \mu(G(f)_{P \times Q}^-)$  is small if and only if  $\sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i$  is small. In other words,  $G(f)$  is Jordan measurable if and only if  $f$  is Riemann integrable. Moreover,  $\mu(G(f))$  is the Riemann integral  $\int_a^b f(x) dx$ .

**Exercise 11.44.** Prove that Jordan measurability implies Lebesgue integrability, and the two measures are equal.

**Exercise 11.45.** Determine the Jordan measurability and compute the volume.

1.  $\{m^{-1} + n^{-1} : m, n \in \mathbb{N}\}$ .
2.  $[0, 1] - \mathbb{Q}$ .
3.  $\{(x, x^2) : 0 \leq x \leq 1\}$ .
4.  $\{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq x^2\}$ .
5.  $\{(x, y) : 0 \leq x \leq y^2, 0 \leq y \leq 1\}$ .
6.  $\{(n^{-1}, y) : n \in \mathbb{N}, 0 < y < 1 + n^{-1}\}$ .
7.  $\{(x, y) : |x| < 1, |y| < 1, x \in \mathbb{Q} \text{ or } y \in \mathbb{Q}\}$ .
8.  $\{(x, y) : |x| < 1, |y| < 1, x \neq y\}$ .

**Exercise 11.46.** Prove that any straight line segment in  $\mathbb{R}^n$  is Jordan measurable, and the volume is 0 when  $n > 1$ . Extend to plane region in  $\mathbb{R}^n$  for  $n > 2$ , etc.

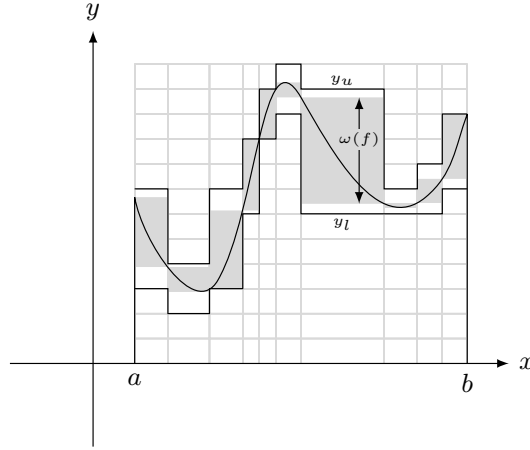


Figure 11.5.1. approximating volume of graph

### Criterion for Measurability

The boundary of subset  $G(f)$  in Example 11.5.4 consists of the graph of  $f$  and some straight lines. As a subset, the graph of  $f$  is approximated from outside by  $G(f)_{P \times Q}^+ - G(f)_{P \times Q}^-$ . The definition of Jordan measure says that  $G(f)$  is Jordan measurable if and only if  $\mu(G(f)_{P \times Q}^+ - G(f)_{P \times Q}^-)$  is as small as we wish. In other words, the volume of the graph of  $f$  is zero, or the volume of the boundary of  $G(f)$  should be zero.

In general,  $\vec{x}$  is a *boundary point* of  $A \subset \mathbb{R}^n$  for any  $\epsilon > 0$ , there are  $\vec{a} \in A$  and  $\vec{b} \notin A$ , such that  $\|\vec{x} - \vec{a}\| < \epsilon$  and  $\|\vec{x} - \vec{b}\| < \epsilon$ . In other words, near  $\vec{x}$  we can find points inside as well as outside  $A$ . The boundary is denoted by  $\partial A$ .

**Proposition 11.5.2.** *The following are equivalent for a bounded  $A \subset \mathbb{R}^n$ .*

1.  $A$  is Jordan measurable: For any  $\epsilon > 0$ , there is a partition  $P$ , such that  $\mu(A_P^+) - \mu(A_P^-) < \epsilon$ .
2. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$  implies  $\mu(A_P^+) - \mu(A_P^-) < \epsilon$ .
3.  $\partial A$  has zero volume.

*Proof.* The second statement is stronger than the first.

It can be shown that  $\partial A$  is always contained in the (closed) rectangles between  $A_P^+$  and  $A_P^-$ . Therefore the first statement implies the third.

Given the third statement, we have  $\partial A \subset I_1 \cup I_2 \cup \dots \cup I_k$  for some rectangles satisfying  $\mu(I_1) + \mu(I_2) + \dots + \mu(I_k) < \epsilon$ . By enlarging each coordinate interval from  $[a, b]$  to  $[a - \delta, b + \delta]$  for some tiny  $\delta > 0$ , we get slightly bigger rectangles  $I_i^\delta$  that still satisfy  $\mu(I_1^\delta) + \mu(I_2^\delta) + \dots + \mu(I_k^\delta) < \epsilon$ . Then it can be shown that, for any partition satisfying  $\|P\| < \delta$ , the rectangles between  $A_P^-$  and  $A_P^+$  must be contained in  $\cup I_i^\delta$ . See Exercise 11.47. This implies  $\mu(A_P^+) - \mu(A_P^-) \leq \mu(\cup I_i^\delta) \leq \sum \mu(I_i^\delta) < \epsilon$ .  $\square$

The boundaries of  $A \cup B$ ,  $A \cap B$  and  $A - B$  are all contained in  $\partial A \cup \partial B$ . Then Proposition 11.5.2 implies the measurable part of the following result.

**Proposition 11.5.3.** *If  $A$  and  $B$  are Jordan measurable, then  $A \cup B$ ,  $A \cap B$  and  $A - B$  are Jordan measurable and*

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B). \quad (11.5.4)$$

The equality is the (finite) additivity and can be reduced to the case  $A$  and  $B$  are disjoint. For disjoint  $A$  and  $B$ , we have  $\mu^-(A) + \mu^-(B) \leq \mu^-(A \cup B)$  and  $\mu^+(A) + \mu^+(B) \geq \mu^+(A \cup B)$ . This implies  $\mu(A \cup B) = \mu(A) + \mu(B)$  when  $A$  and  $B$  are measurable.

Example 11.5.3 shows that the Jordan measure is not countably additive. The problem here is the measurability. If a countable union is measurable, we still have the countably additivity by identifying the Jordan measure with the Lebesgue measure.

To allow more flexibility, define a *general partition* of a subset  $B$  to be  $B = \cup_{I \in P} I$  for Jordan measurable subsets  $I$  satisfying  $\mu(I \cap J) = 0$  for  $I \neq J$  in  $P$ . The subset  $B$  is also Jordan measurable and  $\mu(B) = \sum_{I \in P} \mu(I)$ . The size of the partition is

$$\|P\| = \max\{d(I) : I \in P\}, \quad d(I) = \sup\{\|\vec{x} - \vec{y}\|_\infty : \vec{x}, \vec{y} \in I\}.$$

For a subset  $A \subset B$ , the inner and outer approximations  $A_P^-$ ,  $A_P^+$  may also be defined by (11.5.1). Proposition 11.5.2 can be extended and proved the same as before.

**Proposition 11.5.4.** *Suppose  $B$  is Jordan measurable and  $A \subset B$  is a subset. The following are equivalent.*

1.  $A$  Jordan measurable.
2. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$  for a general partition  $P$  implies  $\mu(A_P^+) - \mu(A_P^-) < \epsilon$ .
3. For any  $\epsilon > 0$ , there is a general partition  $P$ , such that  $\mu(A_P^+) - \mu(A_P^-) < \epsilon$ .
4. For any  $\epsilon > 0$ , there are Jordan measurable  $A^+$  and  $A^-$ , such that  $A^- \subset A \subset A^+$  and  $\mu(A^+) - \mu(A^-) < \epsilon$ .

**Exercise 11.47.** Suppose  $I$  is a subset containing a point  $\vec{a} \in A$  and a point  $\vec{b} \in \mathbb{R}^n - A$ . Prove that there is a point  $\vec{c} \in \partial A$  on the line segment connecting  $\vec{a}$  and  $\vec{b}$ . Moreover, prove that if  $I$  is contained in a ball of radius  $r$ , then  $I \subset B(\vec{c}, 4r)$ . If  $I$  is a rectangle (or a convex set), then  $4r$  can be improved to  $2r$ .

**Exercise 11.48.** Suppose  $A$  and  $B$  differ by boundary:  $A - \partial A \subset B \subset A \cup \partial A$ . Prove that if  $A$  is Jordan measurable, then  $B$  also Jordan measurable and  $\mu(B) = \mu(A)$ . In particular, the *interior*  $A - \partial A$  and the *closure*  $A \cup \partial A$  have the same volume as  $A$ . Conversely,

construct a subset  $A$  such that both the interior and the closure are Jordan measurable, but  $A$  is not Jordan measurable.

**Exercise 11.49.** Suppose  $A \subset \mathbb{R}^m$  and  $B \subset \mathbb{R}^n$  are bounded.

1. Prove that  $\mu^+(A \times B) = \mu^+(A)\mu^+(B)$ .
2. Prove that if  $A$  has volume 0, then  $A \times B$  has volume 0.
3. Prove that if  $\mu^+(A) \neq 0$  and  $\mu^+(B) \neq 0$ , then  $A \times B$  is Jordan measurable if and only if  $A$  and  $B$  are Jordan measurable.
4. Prove that if  $A$  and  $B$  are Jordan measurable, then  $\mu(A \times B) = \mu(A)\mu(B)$ .

**Exercise 11.50.** Prove that  $\mu^+(\partial A) = \mu^+(A) - \mu^-(A)$ . This gives another proof that  $A$  is Jordan measurable if and only if  $\partial A$  has volume 0.

**Exercise 11.51.** Prove that the disks in  $\mathbb{R}^2$  are Jordan measurable by comparing the inscribed and circumscribed regular  $n$ -gons.

## Riemann Sum

Let  $f(\vec{x})$  be a bounded function on a Jordan measurable subset  $A \subset \mathbb{R}^n$ . For any general partition  $P$  of  $A$  and choices  $\vec{x}_I^* \in I$ , define the *Riemann sum*

$$S(P, f) = \sum_{I \in P} f(\vec{x}_I^*)\mu(I). \quad (11.5.5)$$

**Definition 11.5.5.** A bounded function  $f(\vec{x})$  is *Riemann integrable* on a Jordan measurable subset  $A$ , with *Riemann integral*  $J$ , if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\|P\| < \delta \implies |S(P, f) - J| < \epsilon.$$

Like the single variable case, define the *oscillation*

$$\omega_A(f) = \sup_{\vec{x} \in A} f(\vec{x}) - \inf_{\vec{x} \in A} f(\vec{x}). \quad (11.5.6)$$

Moreover, as suggested by Example 11.5.4, define the region between the graph of the function and the  $\vec{x}$ -plane ( $y \in [f(\vec{x}), 0]$  if  $f(\vec{x}) \leq 0$ )

$$G_A(f) = \{(\vec{x}, y) : \vec{x} \in A, y \in [0, f(\vec{x})]\} \subset \mathbb{R}^{n+1}. \quad (11.5.7)$$

**Proposition 11.5.6.** Suppose  $f$  is a bounded function on a Jordan measurable subset. Then the following are equivalent.

1.  $f$  is Riemann integrable.
2. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$  implies  $\sum_{I \in P} \omega_I(f)\mu(I) < \epsilon$ .
3. For any  $\epsilon > 0$ , there is a (general) partition  $P$ , such that  $\sum_{I \in P} \omega_I(f)\mu(I) < \epsilon$ .



4.  $G_A(f)$  is Jordan measurable.

Moreover, if  $f \geq 0$ , then the Riemann integral of  $f$  on  $A$  is the  $(n+1)$ -dimensional volume of  $G_A(f)$ .

*Proof.* The reason for the first two statements to be equivalent is the same as the single variable case. The third statement is a special case of the second.

Assume the third statement. Then the total volume of the “oscillation rectangles”

$$R = \cup [x_{i-1}, x_i] \times \left[ \inf_{[x_{i-1}, x_i]} f, \sup_{[x_{i-1}, x_i]} f \right]$$

is  $< \epsilon$ . Moreover, the boundary of  $G_A(f)$  is contained in the union of  $R$ ,  $A \times 0$ , and  $\partial A \times [-b, b]$ , where  $b$  is a bound for  $|f|$ . By Proposition 11.5.2,  $\partial A$  has volume 0. Then by Exercise 11.49,  $A \times 0$  and  $\partial A \times [-b, b]$  have volume 0. Therefore  $\mu^+(\partial G_A(f)) \leq \mu^+(R) < \epsilon$ . Since  $\epsilon$  can be arbitrarily small,  $\partial G_A(f)$  has volume 0, and we get the fourth statement.

It remains to prove that the fourth statement implies the second. By Proposition 11.5.2, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that for any partition of  $\mathbb{R}^{n+1}$  satisfying  $\|P\| < \delta$ , we have  $\mu(G_A(f)_P^+) - \mu(G_A(f)_P^-) < \epsilon$ . Then by taking  $P$  to be the product of a partition of  $A$  with a partition of  $\mathbb{R}$ , an argument similar to Example 11.5.4 proves the second statement. The argument also shows that, if  $f \geq 0$ , then the Riemann sum of  $f$  on  $A$  is arbitrarily close to the volume of  $G_A(f)$ .  $\square$

The Proposition relates the Riemann integral on  $\mathbb{R}^n$  to the Jordan measure on  $\mathbb{R}^{n+1}$ . Therefore if we treat the volume theory of the Euclidean space of all dimensions at the same time, then the integration theory is equivalent to the measure theory.

For example, let the “upper half” of  $\mathbb{R}^{n+1}$  to be  $H^+ = \{(\vec{x}, y) : y \geq 0\} = \mathbb{R}^n \times [0, +\infty)$ . The similar “lower half” is  $H^-$ . Then  $G_A(f) = (G_A(f) \cap H^+) \cup (G_A(f) \cap H^-)$  is Jordan measurable if and only if

$$G_A(f) \cap H^+ = G_A(\max\{f, 0\}), \quad G_A(f) \cap H^- = G_A(\min\{f, 0\}),$$

are Jordan measurable. This means that  $f$  is Riemann integrable if and only if  $\max\{f, 0\}$  and  $\min\{f, 0\}$  are Riemann integrable, and (see Proposition 11.5.10)

$$\begin{aligned} \int_A f d\mu &= \int_A \max\{f, 0\} d\mu - \int_A \min\{f, 0\} d\mu \\ &= \mu(G_A(f) \cap H^+) - \mu(G_A(f) \cap H^-). \end{aligned}$$

Combined with Proposition 11.5.3, we get the following.

**Proposition 11.5.7.** *Suppose  $f$  is Riemann integrable on Jordan measurable subsets  $A$  and  $B$ . Then  $f$  is Riemann integrable on  $A \cup B$  and  $A \cap B$ , with*

$$\int_{A \cup B} f d\mu + \int_{A \cap B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

**Example 11.5.5.** For the characteristic function  $\chi_A$  of a subset  $A$  contained in a big rectangle  $I$ , we have  $G_I(\chi_A) = A \times [0, 1] \cup I \times 0$ . So  $\chi_A$  is Riemann integrable if and only if  $A \times [0, 1]$  is Jordan measurable, which by Exercise 11.49 is equivalent to that  $A$  is Jordan measurable. The Riemann integral is equal to  $\mu_{n+1}(A \times [0, 1]) = \mu_n(A)$ .

**Exercise 11.52.** Prove that Riemann integrability implies Lebesgue integrability, and the two integrals are equal.

**Exercise 11.53.** Prove that if  $\mu(A) = 0$ , then any bounded function is Riemann integrable on  $A$ , with  $\int_A f d\mu = 0$ .

**Exercise 11.54.** Prove that if  $\int_A f d\mu > 0$ , then there is  $\epsilon > 0$  and Jordan measurable subset  $B \subset A$  with  $\mu(B) > 0$ , such that  $f > \epsilon$  on  $B$ .

**Exercise 11.55.** The Thomae function in Examples 2.3.2 and 4.1.5 may be extended to two variables by  $R(x, y) = \frac{1}{q_x} + \frac{1}{q_y}$  if  $x$  and  $y$  are rational numbers with irreducible denominators  $q_x$  and  $q_y$  and  $R(x, y) = 0$  otherwise. Prove that  $R(x, y)$  is Riemann integrable on any Jordan measurable subset, and the integral is 0.

**Exercise 11.56.** Suppose  $f$  is a bounded function on a Jordan measurable subset  $A$ . Prove that  $f$  is Riemann integrable if and only if  $\hat{G}_A(f) = \{(\vec{x}, y) : \vec{x} \in A, y \in [0, f(\vec{x})]\}$  is Jordan measurable. Moreover, the Riemann integrability of  $f$  implies  $H_A(f) = \{(\vec{x}, f(\vec{x})) : \vec{x} \in A\}$  has volume zero, but the converse is not true.

**Exercise 11.57.** Suppose  $f \leq g \leq h$  on a Jordan measurable subset  $A$ . Prove that if  $f$  and  $h$  are Riemann integrable, with  $\int_A f d\mu = \int_A h d\mu$ , then  $g$  is also Riemann integrable.

**Exercise 11.58.** Prove that if  $f$  is Riemann integrable on Jordan measurable subset  $A$ , then  $f$  is Riemann integrable on any Jordan measurable subset contained in  $A$ .

**Exercise 11.59.** Extend the Riemann integral to a map  $F: A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  on a Jordan measurable subset  $A$ . Define the oscillation  $\omega_A(F) = \sup_{\vec{x}, \vec{y} \in A} \|F(\vec{x}) - F(\vec{y})\|$ .

1. Prove that  $F$  is Riemann integrable if and only if its coordinate functions are Riemann integrable, and the coordinates of the Riemann integral of  $F$  are the Riemann integrals of the coordinates of  $F$ .
2. Prove that  $F$  is Riemann integrable if and only if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$  implies  $\sum_{I \in P} \omega_I(F) \mu(I) < \epsilon$ .
3. Prove that  $F$  is Riemann integrable if and only if for any  $\epsilon > 0$ , there is  $P$ , such that  $\sum_{I \in P} \omega_I(F) \mu(I) < \epsilon$ .
4. Prove  $\left\| \int_A F(\vec{x}) d\mu \right\| \leq \int_A \|F(\vec{x})\| d\mu \leq \left( \sup_{\vec{x} \in A} \|F(\vec{x})\| \right) \mu(A)$ .

Proposition 4.1.4 may be extended.

**Proposition 11.5.8.** *Bounded and continuous maps are Riemann integrable on Jordan measurable subsets.*

Like the earlier result, the proof is based on the uniform continuity. Although  $A$  may not be compact,  $A$  is approximated by  $A_P^-$  from inside, which is compact. Then adding estimations on  $A_P^-$  and on  $A - A_P^-$  together verifies the Riemann integrability.

Proposition 4.1.5 cannot be extended in general because there is no monotonicity concept for multivariable functions. With basically the same proof, Proposition 4.1.6 can be extended.

**Proposition 11.5.9.** *Suppose  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a Riemann integrable map on a Jordan measurable subset  $A$ . Suppose  $\Phi$  is a uniformly continuous map on the values  $F(A)$  of  $F$ . Then the composition  $\Phi \circ F$  is Riemann integrable on  $A$ .*

The Riemann integrability of  $F$  means the integrability of coordinate functions. See Exercise 11.59.

Propositions 4.3.1 and 4.3.2 can also be extended, by the same argument.

**Proposition 11.5.10.** *Suppose  $f$  and  $g$  are Riemann integrable on a Jordan measurable subset  $A$ .*

1. *The linear combination  $af + bg$  and the product  $fg$  are Riemann integrable on  $A$ , and  $\int_A (af + bg)d\mu = a \int_A f d\mu + b \int_A g d\mu$ .*

2. *If  $f \leq g$ , then  $\int_A f d\mu \leq \int_A g d\mu$ . Moreover,  $\left| \int_A f d\mu \right| \leq \int_A |f| d\mu$ .*

## Fubini Theorem

**Theorem 11.5.11 (Fubini Theorem).** *Suppose  $f(\vec{x}, \vec{y})$  is Riemann integrable on  $A \times B$ , where  $A$  and  $B$  are Jordan measurable. Suppose for each fixed  $\vec{y}$ ,  $f(\vec{x}, \vec{y})$  is Riemann integrable on  $A$ . Then  $\int_A f(\vec{x}, \vec{y}) d\mu_{\vec{x}}$  is Riemann integrable on  $B$ , and*

$$\int_{A \times B} f(\vec{x}, \vec{y}) d\mu_{\vec{x}, \vec{y}} = \int_B \left( \int_A f(\vec{x}, \vec{y}) d\mu_{\vec{x}} \right) d\mu_{\vec{y}}.$$

The equality of the Fubini theorem follows from the Lebesgue integral. The issue here is really about integrability.

*Proof.* Let  $g(\vec{y}) = \int_A f(\vec{x}, \vec{y}) d\mu_{\vec{x}}$ . For partitions  $P$  and  $Q$  of  $A$  and  $B$ , we have

$$\begin{aligned} S(P \times Q, f) &= \sum_{I \in P, J \in Q} f(\vec{x}_I^*, \vec{y}_J^*) \mu_{\vec{x}}(I) \mu_{\vec{y}}(J) = \sum_{J \in Q} S(P, f(\vec{x}, \vec{y}_J^*)) \mu_{\vec{y}}(J), \\ S(Q, g) &= \sum_{J \in Q} g(\vec{y}_J^*) \mu_{\vec{y}}(J) = \sum_{J \in Q} \left( \int_A f(\vec{x}, \vec{y}_J^*) d\mu_{\vec{x}} \right) \mu_{\vec{y}}(J). \end{aligned}$$

Since  $f$  is Riemann integrable on  $A \times B$ , for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$ ,  $\|Q\| < \delta$  implies

$$\left| S(P \times Q, f) - \int_{A \times B} f(\vec{x}, \vec{y}) d\mu_{\vec{x}, \vec{y}} \right| < \epsilon.$$

Fix one partition  $Q$  satisfying  $\|Q\| < \delta$  and fix one choice of  $\vec{y}_J^*$ . Then there is  $\delta \geq \delta' > 0$ , such that  $\|P\| < \delta'$  implies

$$\left| S(P, f(\vec{x}, \vec{y}_J^*)) - \int_A f(\vec{x}, \vec{y}_J^*) d\mu_{\vec{x}} \right| < \epsilon$$

for all (finitely many)  $J \in Q$ . This further implies

$$\begin{aligned} |S(P \times Q, f) - S(Q, g)| &\leq \sum_{J \in Q} \left| S(P, f(\vec{x}, \vec{y}_J^*)) - \int_A f(\vec{x}, \vec{y}_J^*) d\mu_{\vec{x}} \right| \mu_{\vec{y}}(J) \\ &\leq \sum_{J \in Q} \epsilon \mu_{\vec{y}}(J) = \epsilon \mu_{\vec{y}}(B). \end{aligned}$$

Then for  $\|P\| < \delta'$ ,  $\|Q\| < \delta'$ , we have

$$\begin{aligned} &\left| S(Q, g) - \int_{A \times B} f(\vec{x}, \vec{y}) d\mu_{\vec{x}, \vec{y}} \right| \\ &\leq |S(P \times Q, f) - S(Q, g)| + \left| S(P \times Q, f) - \int_{A \times B} f(\vec{x}, \vec{y}) d\mu_{\vec{x}, \vec{y}} \right| \\ &< \epsilon \mu_{\vec{y}}(B) + \epsilon. \end{aligned}$$

Therefore  $g$  is Riemann integrable, and  $\int_B g d\mu_{\vec{y}} = \int_{A \times B} f(\vec{x}, \vec{y}) d\mu_{\vec{x}, \vec{y}}$ .  $\square$

**Example 11.5.6.** The two variable Thomae function in Exercise 11.55 is Riemann integrable. However, for each rational  $y$ , we have  $R(x, y) \geq \frac{1}{q_y} > 0$  for rational  $x$  and  $R(x, y) = 0$  for irrational  $x$ . By the reason similar to the non-integrability of the Dirichlet function,  $f(x, y)$  is not integrable in  $x$ . Thus the repeated integral  $\int \left( \int R(x, y) dx \right) dy$  does not exist. The other repeated integral also does not exist for the same reason.

Note that since Riemann integrability is not changed if the function is modified on a subset of volume 0, Fubini Theorem essentially holds even if those  $\vec{y}$  for which  $f(\vec{x}, \vec{y})$  is not Riemann integrable in  $\vec{x}$  form a subset of volume zero. Unfortunately, this is not the case for the two variable Thomae function, because the subset of rational numbers does not have volume zero.

**Example 11.5.7.** If  $f(\vec{x}, \vec{y})$  is Riemann integrable on  $A \times B$ ,  $f(\vec{x}, \vec{y})$  is Riemann integrable on  $A$  for each fixed  $\vec{y}$ , and  $f(\vec{x}, \vec{y})$  is Riemann integrable on  $B$  for each fixed  $\vec{x}$ , then by Fubini Theorem, the two repeated integrals are the same

$$\int_A \left( \int_B f(\vec{x}, \vec{y}) d\mu_{\vec{y}} \right) d\mu_{\vec{x}} = \int_B \left( \int_A f(\vec{x}, \vec{y}) d\mu_{\vec{x}} \right) d\mu_{\vec{y}}.$$

However, if two repeated integrals exist and are equal, it does not necessarily follow that the function is Riemann integrable.

Consider

$$S = \left\{ \left( \frac{k}{p}, \frac{l}{p} \right) : 0 \leq k, l \leq p, p \text{ is a prime number} \right\}.$$

The section  $S_x = \{y : (x, y) \in S\}$  is empty for any irrational  $x$  and is finite for any rational  $x$ . As a result, the section has volume 0 for any  $x$ . The same holds for the sections  $S_y$ . On the other hand, for any rectangular partition  $P$  of  $[0, 1]^2$ , we have  $S_P^+ = [0, 1]^2$  because  $S$  is dense in  $[0, 1]^2$ , and  $S_P^- = \emptyset$  because  $S$  contains no rectangles. Therefore  $\mu(S_P^+) = 1$ ,  $\mu(S_P^-) = 0$ , and  $S$  is not Jordan measurable.

By Example 11.5.5, a subset is Jordan measurable if and only if its characteristic function is Riemann integrable. Therefore in terms of  $\chi_S$ , the two repeated integrals  $\int \left( \int \chi_S(x, y) dy \right) dx$  and  $\int \left( \int \chi_S(x, y) dx \right) dy$  exist and are equal to 0, but the double integral  $\int \chi_S(x, y) dx dy$  does not exist.

**Exercise 11.60.** Study the existence and the equalities between the double integral and the repeated integrals.

1.  $f(x, y) = \begin{cases} 1, & \text{if } x \text{ is rational,} \\ 2y, & \text{if } x \text{ is irrational,} \end{cases} \text{ on } [0, 1] \times [0, 1].$
2.  $f(x, y) = \begin{cases} 1, & \text{if } x \text{ is rational,} \\ 2y, & \text{if } x \text{ is irrational,} \end{cases} \text{ on } [0, 1] \times [0, 2].$
3.  $f(x, y) = \begin{cases} 1, & \text{if } x, y \text{ are rational,} \\ 0, & \text{otherwise.} \end{cases}$
4.  $f(x, y) = \begin{cases} R(x), & \text{if } x, y \text{ are rational,} \\ 0, & \text{otherwise,} \end{cases} \text{ where } R(x) \text{ is the Thomae function in Example 2.3.2.}$
5.  $f(x, y) = \begin{cases} 1, & \text{if } x = \frac{1}{n}, n \in \mathbb{N} \text{ and } y \text{ is rational,} \\ 0, & \text{otherwise.} \end{cases}$
6.  $f(x, y) = \begin{cases} 1, & \text{if } x = \frac{1}{n}, n \in \mathbb{N} \text{ and } y \text{ is rational,} \\ 1, & \text{if } x \text{ is rational and } y = \frac{1}{n}, n \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$

**Exercise 11.61.** Suppose  $f(\vec{x})$  and  $g(\vec{y})$  are bounded functions on Jordan measurable  $A$  and  $B$ .

1. Prove that if  $f(\vec{x})$  and  $g(\vec{y})$  are Riemann integrable, then  $f(\vec{x})g(\vec{y})$  is Riemann integrable, and  $\int_{A \times B} f(\vec{x})g(\vec{y})d\mu_{\vec{x}, \vec{y}} = \int_A f(\vec{x})d\mu_{\vec{x}} \int_B g(\vec{y})d\mu_{\vec{y}}$ .
2. Prove that if  $f(\vec{x})$  is Riemann integrable and  $\int_A f d\mu \neq 0$ , then  $f(\vec{x})g(\vec{y})$  is Riemann integrable if and only if  $g(\vec{y})$  is Riemann integrable.

**Exercise 11.62.** Directly prove the measure version of the Fubini Theorem: If  $A \subset \mathbb{R}^m \times \mathbb{R}^n$  is Jordan measurable, such that the section  $A_{\vec{x}} = \{\vec{y}: (\vec{x}, \vec{y}) \in A\}$  is Jordan measurable for each  $\vec{x}$ , then  $\mu(A_{\vec{x}})$  is Riemann integrable, and  $\mu(A) = \int \mu(A_{\vec{x}})d\mu_{\vec{x}}$ .

1. Let  $P$  and  $Q$  be a partitions of  $\mathbb{R}^m$  and  $\mathbb{R}^n$ . Prove that  $\vec{x} \in I \in P$  implies  $\cup_{J \in Q, I \times J \subset A} J \subset A_{\vec{x}} \subset \cup_{J \in Q, I \times J \cap A \neq \emptyset} J$ .
2. Choose  $\vec{x}_I^* \in I$  for every  $I \in P$ , such that  $A_{\vec{x}_I^*}$  are all Jordan measurable. Prove that  $|S(P, \mu_n(A_{\vec{x}})) - \mu_{m+n}(A)| \leq \mu_{m+n}(A_{P \times Q}^+) - \mu_{m+n}(A_{P \times Q}^-)$ .
3. Prove that if  $f(\vec{x})$  is Riemann integrable and  $\int_A f d\mu \neq 0$ , then  $f(\vec{x})g(\vec{y})$  is Riemann integrable if and only if  $g(\vec{y})$  is Riemann integrable.

**Exercise 11.63.** By using the equivalence between the Riemann integral of  $f$  and the volume of  $G(f)$ , derive the Fubini Theorem for Riemann integrals from the Fubini theorem for the Jordan measure.

## 11.6 Additional Exercise

### Fubini Theorem for Jordan Measurable Subset

In the Lebesgue measure theory, the Fubini Theorem for integrals (Theorem 11.3.4) was derived from the Fubini Theorem for measurable subsets (Proposition 11.3.3). We try to do the same for the Jordan measure.

Let  $A \subset \mathbb{R}^m \times \mathbb{R}^n$  be a subset. For any  $\vec{x} \in \mathbb{R}^m$ , define the section  $A_{\vec{x}} = \{\vec{y}: (\vec{x}, \vec{y}) \in A\}$ . Let  $P$  and  $Q$  be partitions of  $\mathbb{R}^m$  and  $\mathbb{R}^n$ .

**Exercise 11.64.** For  $\vec{x} \in I \in P$ , prove that

$$\cup_{J \in Q, I \times J \subset A} J \subset A_{\vec{x}} \subset \cup_{J \in Q, I \times J \cap A \neq \emptyset} J.$$

**Exercise 11.65.** Prove that

$$I \times (\cup_{J \in Q, I \times J \subset A} J) = A_{P \times Q}^- \cap I \times \mathbb{R}^n, \quad I \times (\cup_{J \in Q, I \times J \cap A \neq \emptyset} J) = A_{P \times Q}^+ \cap I \times \mathbb{R}^n.$$

**Exercise 11.66.** For any choice  $\vec{x}_I^* \in I$  for every  $I \in P$ , prove that

$$A_{P \times Q}^- \subset \cup_{I \in P} I \times A_{\vec{x}_I^*} \subset A_{P \times Q}^+.$$

**Exercise 11.67.** Suppose  $A$  is Jordan measurable. Suppose  $\vec{x}_I^* \in I$  are chosen for every  $I \in P$ , such that  $A_{\vec{x}_I^*}$  are all Jordan measurable. Prove that

$$|S(P, \mu_n(A_{\vec{x}})) - \mu_{m+n}(A)| \leq \mu_{m+n}(A_{P \times Q}^+) - \mu_{m+n}(A_{P \times Q}^-).$$

**Exercise 11.68.** Prove the Jordan measure version of the Fubini Theorem: If  $A \subset \mathbb{R}^m \times \mathbb{R}^n$  is Jordan measurable, such that the section  $A_{\vec{x}}$  is Jordan measurable for each  $\vec{x}$ , then  $\mu(A_{\vec{x}})$  is Riemann integrable, and  $\mu(A) = \int \mu(A_{\vec{x}}) d\mu_{\vec{x}}$ .

**Exercise 11.69.** By using the equivalence between the Riemann integral of  $f$  and the volume of  $G(f)$ , derive the Fubini Theorem for Riemann integrals from the Fubini theorem for the Jordan measure.





## Chapter 12

# Differentiation of Measure

## 12.1 Radon-Nikodym Theorem

For a non-negative integrable function  $f$  on a measure space  $(X, \Sigma, \mu)$ , by Proposition 10.3.9 (also see Exercise 10.39),

$$\nu(A) = \int_A f d\mu \quad (12.1.1)$$

is a measure on the  $\sigma$ -algebra  $\Sigma$ . This suggests that we may sometimes compare measures and regard  $f$  as the “derivative” of  $\nu$  with respect to  $\mu$ .

The consideration does not have to be restricted to non-negative  $f$  only. For general integrable  $f$ , we may get a measure with possibly negative value.

**Definition 12.1.1.** A *signed measure* on a  $\sigma$ -algebra  $\Sigma$  is a function  $\nu$  on  $\Sigma$  satisfying the following properties.

1. *Empty Set:*  $\nu(\emptyset) = 0$ .
2. *Countable Additivity:* If  $A_i \in \Sigma$  are disjoint, then  $\nu(\sqcup A_i) = \sum \nu(A_i)$ .
3. *Infinity:*  $\nu$  cannot take both  $+\infty$  and  $-\infty$  as values.

In the second property, the sum  $\sum \nu(A_i)$  must converge absolutely if  $\nu(\sqcup A_i)$  is finite. It also implies that, if  $A \subset B$  and  $\mu(B)$  is finite, then  $\mu(A)$  is finite.

The third property allows  $\nu$  to have extended value. However, the choice of infinity value is restricted so that the additivity still makes sense. In particular, if  $\nu(\sqcup A_i) = +\infty$ , then  $\sum \nu(A_i)$  diverges to  $+\infty$ . This means that the sum of negative terms converges and the sum of positive terms is  $+\infty$ .

**Exercise 12.1.** Suppose  $\nu_1$  and  $\nu_2$  are signed measures on the same  $\sigma$ -algebra  $\Sigma$  on  $X$ . If both do not take  $-\infty$  value, prove that  $\nu_1 + \nu_2$  is a signed measure. Moreover, if  $\nu_1$  and  $\nu_2$  are  $\sigma$ -finite, then  $\nu_1 + \nu_2$  is  $\sigma$ -finite.

**Exercise 12.2.** Suppose  $f$  is a measurable function. If  $\nu(A) = \int_A f d\mu$  has finite lower bound (but may take  $+\infty$  value), prove that  $\nu$  is a signed measure. Moreover, if  $\mu$  is  $\sigma$ -finite and  $\mu(\{x: f(x) = +\infty\}) = 0$ , prove that  $\nu$  is  $\sigma$ -finite.

### Hahn Decomposition

**Theorem 12.1.2.** Suppose  $\nu$  is a signed measure on a  $\sigma$ -algebra  $\Sigma$  on a set  $X$ . Then there is a disjoint union  $X = X^+ \sqcup X^-$  with  $X^+, X^- \in \Sigma$ , such that

$$\begin{aligned} A \subset X^+, A \in \Sigma &\implies \nu(A) \geq 0, \\ A \subset X^-, A \in \Sigma &\implies \nu(A) \leq 0. \end{aligned}$$

The decomposition  $X = X^+ \sqcup X^-$  in the theorem is called a *Hahn decomposition* of the signed measure.

*Proof.* We call  $Y \in \Sigma$  a *positive subset* if it satisfies

$$A \subset Y, A \in \Sigma \implies \nu(A) \geq 0.$$

The subset  $X^+$  should be in some sense a maximal positive subset.

Suppose  $Z \in \Sigma$  satisfies  $+\infty > \nu(Z) > 0$ . We will find a positive subset  $Y \subset Z$ , by deleting some subsets of negative measure from  $Z$ . Let  $n_1$  be the smallest natural number, such that there is  $Z_1 \in \Sigma$  satisfying

$$Z_1 \subset Z, \quad \nu(Z_1) < -\frac{1}{n_1}.$$

Let  $n_2$  be the smallest natural number, such that there is  $Z_2 \in \Sigma$  satisfying

$$Z_2 \subset Z - Z_1, \quad \nu(Z_2) < -\frac{1}{n_2}.$$

Inductively, we have  $Z_k \in \Sigma$  satisfying

$$Z_k \subset Z - Z_1 \sqcup \cdots \sqcup Z_{k-1}, \quad \nu(Z_k) < -\frac{1}{n_k}.$$

Moreover,  $n_k$  being smallest means that any measurable

$$A \in \Sigma, A \subset Z - Z_1 \sqcup \cdots \sqcup Z_{k-1} \implies \nu(A) \geq -\frac{1}{n_k - 1}.$$

Let  $Y = Z - \sqcup Z_k$ . Then  $\nu(Z) = \nu(Y) + \sum \nu(Z_k)$  by the countable additivity. Since  $\nu(Z)$  is finite,  $\sum \nu(Z_k)$  absolutely converges, and by the comparison test (Proposition 5.2.1),  $\sum \frac{1}{n_k}$  also converges. On the other hand, for any  $k$  we have

$$A \in \Sigma, A \subset Y \implies A \subset Z - Z_1 \sqcup \cdots \sqcup Z_{k-1} \implies \nu(A) \geq -\frac{1}{n_k - 1}.$$

The convergence of  $\sum \frac{1}{n_k}$  implies  $\lim_{k \rightarrow \infty} \frac{1}{n_k - 1} = 0$ , and the above further implies  $\nu(A) \geq 0$ . This proves that  $Y$  is a positive subset, and  $\nu(Y) \geq \nu(Z) > 0$ .

Suppose  $\nu$  does not take  $+\infty$  as value. Let

$$b = \sup\{\nu(Y) : Y \text{ is a positive subset}\}.$$

If  $b = 0$ , then the theorem holds for  $X^+ = \emptyset$  and  $X^- = X$ . If  $b > 0$ , then we have  $\lim \nu(Y_k) = b$  for a sequence of positive subsets  $Y_k$ . By the countable additivity, it is easy to see that  $X^+ = \cup Y_k$  is a positive subset satisfying  $\nu(X^+) \geq \nu(Y_k)$ . This implies  $\nu(X^+) = b$ . Now we need to show that  $X^- = X - X^+$  contains no subsets of positive measure. If there is  $Z \subset X^-$  with  $\nu(Z) > 0$ , then there is a positive subset  $Y \subset Z$  with  $\nu(Y) > 0$ . The union  $X^+ \cup Y$  is still positive, with  $\nu(X^+ \cup Y) = \nu(X^+) + \nu(Y) > \nu(X^+) = b$ . This contradicts with the definition of  $b$ . Therefore we conclude that  $X^-$  contains no subset of positive measure.

If  $\nu$  does not take  $-\infty$  as value, then  $-\nu$  does not take  $+\infty$  as value. The decomposition for  $-\nu$  is also the decomposition for  $\nu$ , with the positive and negative parts exchanged.  $\square$

**Exercise 12.3.** Prove the uniqueness of Hahn decomposition: If two decompositions  $X = X^+ \sqcup X^- = Y^+ \sqcup Y^-$  satisfy the properties of Theorem 12.1.2, then any measurable subset of  $X^+ \cap Y^- = X^+ - Y^+ = Y^- - X^-$  or  $X^- \cap Y^+ = X^- - Y^- = Y^+ - X^+$  has zero measure.

**Exercise 12.4.** Suppose two signed measures  $\nu$  and  $\nu'$  satisfy  $\nu(A) \geq \nu'(A)$  for any measurable  $A$ . Will  $\nu$  have bigger  $X^+$  and smaller  $X^-$  in the Hahn decomposition (up to subsets of measure zero as described in Exercise 12.3)?

### Jordan Decomposition

From the Hahn decomposition, we get a decomposition of the signed measure

$$\nu = \nu^+ - \nu^-, \quad \nu^+(A) = \nu(A \cap X^+), \quad \nu^-(A) = -\nu(A \cap X^-).$$

The measures  $\nu^+$  and  $\nu^-$  are independent in the following strong sense.

**Definition 12.1.3.** Two (signed) measures  $\nu$  and  $\nu'$  are *mutually singular*, denoted  $\nu \perp \nu'$ , if there is a measurable decomposition  $X = Y \sqcup Y'$ , such that  $\nu(A) = 0$  for any measurable  $A \subset Y'$  and  $\nu'(A) = 0$  for any measurable  $A \subset Y$ .

Exercise 12.5 shows that the decomposition of a signed measure into the difference of mutually singular measures is unique. We call  $\nu = \nu^+ - \nu^-$  the *Jordan decomposition*, and may further define the *absolute value* of signed measure

$$|\nu|(A) = \nu^+(A) + \nu^-(A) = \nu(A \cap X^+) - \nu(A \cap X^-).$$

**Example 12.1.1.** If  $\nu$  is given by (12.1.1), then the subset  $X^+$  in the Hahn decomposition can be any measurable subset between  $f^{-1}[0, +\infty)$  and  $f^{-1}(0, +\infty)$ . Its complement  $X^-$  can be any measurable subset between  $f^{-1}(-\infty, 0)$  and  $f^{-1}(-\infty, 0]$ . Correspondingly, the Jordan decomposition is given by

$$\begin{aligned} \nu^+(A) &= \int_{A \cap f^{-1}[0, +\infty)} f d\mu = \int_{A \cap f^{-1}(0, +\infty)} f d\mu = \int_A f^+ d\mu, \\ \nu^-(A) &= \int_{A \cap f^{-1}(-\infty, 0]} f d\mu = \int_{A \cap f^{-1}(-\infty, 0)} f d\mu = \int_A f^- d\mu. \end{aligned}$$

The absolute value is

$$|\nu|(A) = \int_A |f| d\mu.$$

**Example 12.1.2.** Consider the Lebesgue-Stieltjes measure  $\mu_\kappa$  induced by the Cantor function  $\kappa$  in Example 11.4.5. Since  $\kappa$  is constant on any interval  $I \subset [0, 1] - K$ , we have  $\mu_\kappa(I) = 0$ . Since  $[0, 1] - K$  is a countable union of such intervals, we get  $\mu_\kappa([0, 1] - K) = 0$ . On the other hand, we have  $\mu(K) = 0$ . Therefore  $[0, 1] = K \sqcup ([0, 1] - K)$  is a decomposition that gives the mutually singular relation  $\mu_\kappa \perp \mu$ .

**Exercise 12.5.** Suppose  $\nu$  is a signed measure, with Hahn decomposition  $X = X^+ \sqcup X^-$ . Suppose  $\nu = \nu^+ - \nu^-$ , and  $\nu^+, \nu^-$  are mutually singular with respect to  $X = Y^+ \sqcup Y^-$ .

1. Prove that  $\nu(A) = \nu^+(A) = \nu^-(A) = 0$  for any  $A \subset X^+ \cap Y^-$  as well as any  $A \subset X^- \cap Y^+$ .
2. Prove that  $\nu^+(A) = \nu(A \cap X^+)$  and  $\nu^-(A) = \nu(A \cap X^-)$ .

This proves the uniqueness of Jordan decomposition.

**Exercise 12.6.** Suppose  $\nu$  is a signed measure, with Hahn decomposition  $X = X^+ \sqcup X^-$ . Suppose  $\nu = \mu^+ - \mu^-$  for measures  $\mu^+$  and  $\mu^-$ .

1. Prove that  $\nu^+(A) \leq \mu^+(A)$  for  $A \subset X^+$  and  $\nu^-(A) \leq \mu^-(A)$  for  $A \subset X^-$ .
2. Prove that  $|\nu|(A) \leq \mu^+(A) + \mu^-(A)$  for general measurable  $A$ .
3. Prove that if  $|\nu|(A) = \mu^+(A) + \mu^-(A) < \infty$  for some measurable  $A$ , then  $\nu^+(B) = \mu^+(B)$  and  $\nu^-(B) = \mu^-(B)$  for any measurable  $B \subset A$ .

This shows that the Jordan decomposition is the “most efficient” in the sense of Proposition 4.6.2.

**Exercise 12.7.** Suppose  $\mu$  is a semi-finite measure. Prove that  $\nu(A) = \int_A f d\mu$  and  $\nu'(A) = \int_A g d\mu$  are mutually singular if and only if  $fg = 0$  almost everywhere.

## Radon-Nikodym Theorem

The integral of a measurable function gives a signed measure (12.1.1). Conversely, we would like to ask whether a signed measure is given by an integral. The following is a necessary condition.

**Definition 12.1.4.** A signed measure  $\nu$  is *absolutely continuous* with respect to a measure  $\mu$ , denoted  $\nu \ll \mu$ , if

$$\mu(A) = 0 \implies \nu(A) = 0.$$

The simple condition actually implies more in case  $\nu$  is finite.

**Proposition 12.1.5.** *If a signed measure  $\nu$  is absolutely continuous with respect to a measure  $\mu$ , and  $\nu(X) < +\infty$ , then for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\mu(A) < \delta$  implies  $|\nu(A)| < \epsilon$ .*

In case  $\mu$  is  $\sigma$ -finite, the proposition is a consequence of Exercise 10.29 and the subsequent Theorem 12.1.6. Our proof of the proposition is more general and does not use the subsequent theorem.

*Proof.* The absolute continuity of  $\nu$  implies the absolute continuity of  $\nu^+(A) = \nu(A \cap X^+)$  and  $\nu^-(A) = \nu(A \cap X^-)$ . So we only need to consider the case  $\nu$  is a measure.

Suppose there is  $\epsilon > 0$ , such that for any  $n$ , there is measurable  $A_n$  satisfying  $\mu(A_n) < \frac{1}{2^n}$  and  $\nu(A_n) \geq \epsilon$ . Then  $B_n = \cup_{i>n} A_i$  is a decreasing sequence satisfying

$\mu(B_n) \leq \sum_{i>n} \mu(A_i) < \frac{1}{2^n}$  and  $\nu(B_n) \geq \nu(A_{n+1}) \geq \epsilon$ . Since  $\nu(X)$  is finite, by the monotone limit property in Proposition 9.4.4, the intersection  $C = \cap B_n$  satisfies  $\nu(C) = \lim_{n \rightarrow \infty} \nu(B_n) \geq \epsilon$ . Moreover,  $\mu(C) \leq \mu(B_n) < \frac{1}{2^n}$  for all  $n$  implies  $\mu(C) = 0$ . This contradicts the absolute continuity of  $\nu$  with respect to  $\mu$ .  $\square$

**Exercise 12.8.** Prove properties of absolutely continuous measures.

1.  $\mu \ll |\mu|$ .
2.  $\nu \ll \mu$  and  $\mu \ll \lambda \implies \nu \ll \lambda$ .
3.  $|\nu| \ll \mu \implies \nu \ll \mu$ .
4.  $\nu \ll \mu \implies \nu \ll |\mu|$ .
5.  $\nu \ll \mu$  and  $\nu' \ll \mu \implies \nu + \nu' \ll \mu$ .

**Exercise 12.9.** Prove properties of mutually singular and absolutely continuous measures.

1.  $\nu \perp \mu$  and  $\nu \ll \mu \implies \nu = 0$ .
2.  $\nu \perp \mu$  and  $\nu' \perp \mu \implies \nu + \nu' \perp \mu$ .
3.  $\nu \perp \mu \iff \nu \perp |\mu| \iff |\nu| \perp |\mu|$ .
4.  $\nu \ll \mu$  and  $\mu \perp \mu' \implies \nu \perp \mu'$ .

**Exercise 12.10.** Show that Proposition 12.1.5 fails if  $\nu(X)$  is not finite. Even  $\sigma$ -finite is not enough.

The absolute continuity turns out to be sufficient for a signed measure to be given by (12.1.1).

**Theorem 12.1.6 (Radon-Nikodym Theorem).** *Suppose  $(X, \Sigma, \mu)$  is a  $\sigma$ -finite measure space. Suppose  $\nu$  is a signed measure on  $\Sigma$ . If  $\nu$  is absolutely continuous with respect to  $\mu$ , then there is a measurable (perhaps extended valued) function  $f$ , such that  $\nu$  is the integral of  $f$  with respect to  $\mu$ .*

Exercise 12.11 shows that the the function  $f$  in the theorem is unique. Exercises 12.36 to 12.39 give another proof of the theorem under stronger assumption.

*Proof.* First assume  $\mu(X)$  is finite. For any number  $a$ , let  $X = X_a^+ \sqcup X_a^-$  be a Hahn decomposition of  $\nu - a\mu$ . The function  $f$  can be constructed by the fact that  $f^{-1}[a, +\infty) \supset X_a^+ \supset f^{-1}(a, +\infty)$ .

For  $a > b$ ,  $X_a^+$  should be almost contained in  $X_b^+$  in the sense that  $X_a^+ - X_b^+$  should have zero measure. In fact, we claim that

$$a > b, A \subset X_a^+ - X_b^+, A \in \Sigma \implies \mu(A) = \nu(A) = 0. \quad (12.1.2)$$

For  $A \subset X_a^+ - X_b^+$ , we have  $(\nu - a\mu)(A) \geq 0$  by  $A \subset X_a^+$  and  $(\nu - b\mu)(A) \leq 0$  by  $A \subset X_b^-$ . Therefore  $a\mu(A) \leq \nu(A) \leq b\mu(A)$ . Since  $a > b$  and  $\mu$  is finite and non-negative, we get  $\mu(A) = \nu(A) = 0$ .

The property (12.1.2) allows us to assume

$$a > b \implies X_a^+ \subset X_b^+. \quad (12.1.3)$$

This can be achieved by replacing  $X_a^+$  with  $\cup_{r \geq a, r \in \mathbb{Q}} X_r^+$ . Strictly speaking, we need to verify that  $\cup_{r \geq a, r \in \mathbb{Q}} X_r^+$  can be used as the positive subset in the Hahn decomposition for  $\nu - a\mu$ . On the one hand, we have

$$\begin{aligned} A \subset X_r^+, r \geq a &\implies A - X_a^+ \subset X_r^+ - X_a^+, r \geq a \\ &\implies (\nu - a\mu)(A - X_a^+) = 0 \\ &\implies (\nu - a\mu)(A) = (\nu - a\mu)(A \cap X_a^+) \geq 0, \end{aligned}$$

where the second implication is by (12.1.2). By countable additivity, we then get

$$A \subset \cup_{r \geq a, r \in \mathbb{Q}} X_r^+ \implies (\nu - a\mu)(A) \geq 0.$$

On the other hand, we have

$$\begin{aligned} A \cap (\cup_{r \geq a, r \in \mathbb{Q}} X_r^+) = \emptyset &\implies A \cap X_r^+ = \emptyset \text{ for any rational } r \geq a \\ &\implies (\nu - r\mu)(A) \leq 0 \text{ for any rational } r \geq a \\ &\implies (\nu - a\mu)(A) \leq 0. \end{aligned}$$

This justifies the use of  $\cup_{r \geq a, r \in \mathbb{Q}} X_r^+$  as the positive subset in the Hahn decomposition for  $\nu - a\mu$ .

Under the assumption (12.1.3), we define

$$f(x) = \sup\{a : x \in X_a^+\}.$$

Then we have

$$\begin{aligned} f(x) > b &\iff x \in X_a^+ \text{ for some } a > b \\ &\iff x \in X_r^+ \text{ for some rational } r > b, \end{aligned}$$

where the second equivalence uses (12.1.3) and may be obtained by choosing any rational  $r$  satisfying  $a > r > b$ . The equivalence means

$$f^{-1}(b, +\infty) = \cup_{r > b, r \in \mathbb{Q}} X_r^+. \quad (12.1.4)$$

In particular,  $f$  is measurable.

There are two technical problems about the definition of  $f$ . The first is that  $f$  would take value  $+\infty$  on  $Y = \cap_{r \in \mathbb{Q}} X_r^+$ . For any measurable  $A \subset Y$ , we have  $\nu(A) \geq r\mu(A)$  for all  $r \in \mathbb{Q}$ . This means that either  $\mu(A) = 0$  or  $\mu(A) > 0$  and  $\nu(A) = +\infty$ .

The second problem is that  $f$  would take value  $-\infty$  on  $Z = \cap_{r \in \mathbb{Q}} X_r^- = X - \cup_{r \in \mathbb{Q}} X_r^+$ . For any measurable  $A \subset Z$ , we have  $\nu(A) \leq r\mu(A)$  for all  $r \in \mathbb{Q}$ . Since  $\mu(A)$  is finite, this means that either  $\mu(A) = 0$  or  $\mu(A) > 0$  and  $\nu(A) = -\infty$ .

The signed measure  $\nu$  cannot take both  $+\infty$  and  $-\infty$  as values. If  $\nu$  does not take  $-\infty$  as value, then we must have  $\mu(Z) = 0$ , and we may take  $f = +\infty$  on  $Y$

and  $f = 0$  on  $Z$ . If  $\nu$  does not take  $+\infty$  as value, then we must have  $\mu(Y) = 0$ , and we may take  $f = 0$  on  $Y$  and  $f = -\infty$  on  $Z$ .

It remains to prove (12.1.1). This follows from the estimation of the  $\nu$ -measure of measurable subsets  $A \subset f^{-1}(a, b]$ . By (12.1.3) and (12.1.4), we have

$$A \subset f^{-1}(a, +\infty) \implies A \subset X_a^+ \implies \nu(A) \geq a\mu(A).$$

By (12.1.4), we also have

$$\begin{aligned} A \cap f^{-1}(b, +\infty) = \emptyset &\implies A \cap X_r^+ = \emptyset \text{ for any rational } r > b \\ &\implies \nu(A) \leq r\mu(A) \text{ for any rational } r > b \\ &\implies \nu(A) \leq b\mu(A). \end{aligned}$$

Combining the two estimations, we get

$$A \subset f^{-1}(a, b] = f^{-1}(a, +\infty) - f^{-1}(b, +\infty) \implies a\mu(A) \leq \nu(A) \leq b\mu(A). \quad (12.1.5)$$

If  $A \in \Sigma$  is disjoint from  $Y$  and  $Z$ , then for any (infinite) partition  $\Pi$  of  $\mathbb{R}$ , we get a partition of  $A$

$$A = \sqcup_{i \in \mathbb{Z}} A_i, \quad A_i = A \cap f^{-1}(c_{i-1}, c_i], \quad (c_{i-1}, c_i] \in \Pi.$$

Since  $c_{i-1} < f \leq c_i$  on  $A_i$ , we have

$$c_{i-1}\mu(A_i) \leq \int_{A_i} f d\mu \leq c_i\mu(A_i).$$

By (12.1.5), we also have

$$c_{i-1}\mu(A_i) \leq \nu(A_i) \leq c_i\mu(A_i).$$

Therefore

$$\left| \nu(A) - \int_A f d\mu \right| \leq \sum \left| \nu(A_i) - \int_{A_i} f d\mu \right| \leq \sum (c_i - c_{i-1})\mu(A_i) \leq \|\Pi\|\mu(A).$$

Since  $\mu(A)$  is finite, this proves (12.1.1) for  $A$  disjoint from  $Y$  and  $Z$ .

Now we verify (12.1.1) for measurable  $A \subset Y$ . The first case is  $\mu(Z) = 0$  and  $f = +\infty$  on  $Y$ . If  $\mu(A) = 0$ , then  $\nu(A) = 0$  by the absolute continuity assumption, and we get  $\nu(A) = 0 = \int_A f d\mu$ . If  $\mu(A) > 0$ , then we know  $\nu(A) = +\infty$ , and we get  $\nu(A) = +\infty = \int_A f d\mu$ . The second case is  $\mu(Y) = 0$  and  $f = 0$  on  $Y$ . Since  $\mu$  is a measure, we get  $\mu(A) = 0$ . By absolute continuity assumption, we have  $\nu(A) = 0$ . Therefore  $\nu(A) = 0 = \int_A f d\mu$ .

The equality (12.1.1) can be similarly verified for measurable  $A \subset Z$ . Then by decomposing  $\nu(A)$  and  $\int_A f d\mu$  according to  $A = (A - Y - Z) \sqcup (A \cap Y) \sqcup (A \cap Z)$ , we prove the equality in general.



We proved the theorem under the assumption that  $\mu(X)$  is finite. For  $\sigma$ -finite  $\mu$ , we have a countable disjoint union  $X = \sqcup X_i$  with all  $\mu(X_i)$  finite. By adding the conclusion of the theorem on each  $X_i$  together, we prove the theorem for the  $\sigma$ -finite  $\mu$ .  $\square$

**Example 12.1.3.** Let  $\mu$  and  $\nu$  be the counting measure and the Lebesgue measure on  $\mathbb{R}$ . Then  $\mu(A) = 0$  implies  $A = \emptyset$ , so that  $\nu(A) = 0$ . Therefore  $\nu$  is absolutely continuous with respect to  $\mu$ . However, by Exercise 10.22, there is no function  $f$  satisfying  $\nu(A) = \int_A f d\mu$ . This shows the necessity of the  $\sigma$ -finite condition in the Radon-Nikodym Theorem.

Suppose  $\mu$  and  $\nu$  only takes finite non-negative values. Then the function  $f$  in the Radon-Nikodym Theorem should be the “upper bound” of the non-negative measurable functions satisfying

$$\nu(A) \geq \int_A f d\mu \text{ for any measurable } A. \quad (12.1.6)$$

The subsequent exercises make use of the idea and give another proof of the Radon-Nikodym Theorem.

**Exercise 12.11.** Suppose  $f$  and  $g$  are extended valued measurable functions, such that  $\int_A f d\mu = \int_A g d\mu$  for any measurable subset  $A$ . Prove that if  $\mu$  is  $\sigma$ -finite, then  $f = g$  almost everywhere. Explain that the  $\sigma$ -finite condition is necessary.

**Exercise 12.12.** Suppose  $f$  is an extended valued measurable function, such that  $\nu(A) = \int_A f d\mu$  is a  $\sigma$ -finite signed measure. Prove that the subset on which  $f = \pm\infty$  has zero measure. This means that we may modify  $f$  not to take any infinity value.

## Differentiation of Measure

**Definition 12.1.7.** Suppose  $\mu$  is a signed measure and  $\nu$  is a measure on the same  $\sigma$ -algebra  $\Sigma$ . If there is a measurable function  $f$ , such that

$$\nu(A) = \int_A f d\mu$$

for any measurable  $A$ , then  $\nu$  is *differentiable* with respect to  $\mu$ , and  $\frac{d\nu}{d\mu} = f$  is the *Radon-Nikodym derivative*.

The Radon-Nikodym Theorem says that, if  $\mu$  is  $\sigma$ -finite, then  $\nu$  is differentiable with respect to  $\mu$  if and only if  $\nu$  is absolutely continuous with respect to  $\mu$ . Example 12.1.3 shows that the Radon-Nikodym derivative for absolutely continuous measures may not exist in general.

**Proposition 12.1.8.** *Suppose a  $\sigma$ -finite measure  $\nu$  is differentiable with respect to a measure  $\mu$ , and  $f = \frac{d\nu}{d\mu}$  is the Radon-Nikodym derivative. Then  $g$  is integrable with respect to  $\nu$  if and only if  $gf$  is integrable with respect to  $\mu$ , and*

$$\int g d\nu = \int gf d\mu.$$

Define the integral with respect to a signed measure  $\nu$  by using the Jordan decomposition

$$\int g d\nu = \int g d\nu^+ - \int g d\nu^-.$$

In particular, the integrability means the integrability of the two integrals on the right, which further means the integrability of all four  $\int g^\pm d\nu^\pm$ . The definition can also be extended to allow  $\pm\infty$  values, and the expressions such as  $(+\infty) - c = +\infty$  or  $(+\infty) - (-\infty) = +\infty$  are allowed on the right. Then Proposition 12.1.8 extends to  $\sigma$ -finite  $\nu$ .

*Proof.* We first consider the case  $\nu(X)$  is finite, and  $g$  does not take  $\pm\infty$  as value. Since  $\nu$  is non-negative, we can also assume  $f \geq 0$ . For any (infinite) partition  $\Pi$  of  $\mathbb{R}$ , because  $g$  does not take  $\pm\infty$ , we get a partition

$$X = \sqcup_{i \in \mathbb{Z}} A_i, \quad A_i = g^{-1}(c_{i-1}, c_i], \quad (c_{i-1}, c_i] \in \Pi.$$

By  $c_{i-1} < g \leq c_i$  on  $A_i$  and  $\nu \geq 0$ , we get

$$c_{i-1}\nu(A_i) \leq \int_{A_i} g d\nu \leq c_i\nu(A_i).$$

By  $f \geq 0$ , we also have

$$c_{i-1}\nu(A_i) \leq \int_{A_i} c_{i-1} f d\mu \leq \int_{A_i} g f d\mu \leq \int_{A_i} c_i f d\mu \leq c_i\nu(A_i).$$

Therefore

$$\left| \int g d\nu - \int g f d\mu \right| \leq \sum \left| \int_{A_i} g d\nu - \int_{A_i} g f d\mu \right| \leq \sum (c_i - c_{i-1})\nu(X) \leq \|\Pi\|\nu(X).$$

Since  $\nu(X)$  is finite, we get  $\int g d\nu = \int g f d\mu$ .

The equality extends to  $\sigma$ -finite  $\nu$ , by expressing the integrals into a sum of integrals on subsets with finite  $\nu$ -measure.

It remains to consider the possibility that  $g$  might take  $+\infty$  value. In other words, assume  $g = +\infty$  on measurable  $A$ . We need to show that  $\int_A g d\nu = \int_A g f d\mu$ .

First consider  $\nu(A) > 0$ . By  $g = +\infty$  on  $A$ , we get  $\int_A g d\nu = +\infty$ . On the other hand, for  $B = \{x \in A: f(x) > 0\}$ , we have  $\int_B f d\mu = \int_A f d\mu = \nu(A) > 0$ .

This implies  $\mu(B) > 0$ . By  $gf = +\infty$  on  $B$  and  $gf = 0$  on  $A - B$ , we get  $\int_A gf d\mu \geq \int_B gf d\mu = +\infty$ . We conclude that  $\int_A g d\nu = +\infty = \int_A gf d\mu$  in case  $\nu(A) > 0$ .

Next consider  $\nu(A) = 0$ . We have  $\int_A g d\nu = 0$ . On the other hand, for  $B = \{x \in A : f(x) > 0\}$  above, we have  $\int_B f d\mu = \int_A f d\mu = \nu(A) = 0$ . This implies  $\mu(B) = 0$ . By  $gf = 0$  on  $A - B$  and  $\mu(B) = 0$ , we get  $\int_A gf d\mu = \int_B gf d\mu = 0$ . We conclude that  $\int_A g d\nu = 0 = \int_A gf d\mu$  in case  $\nu(A) = 0$ .  $\square$

**Exercise 12.13.** Prove the properties of the Radon-Nikodym derivative

$$\frac{d(\lambda + \nu)}{d\mu} = \frac{d\lambda}{d\mu} + \frac{d\nu}{d\mu}, \quad \frac{d(c\nu)}{d\mu} = c \frac{d\nu}{d\mu}, \quad \frac{d|\nu|}{d\mu} = \left| \frac{d\nu}{d\mu} \right|, \quad \frac{d\lambda}{d\mu} = \frac{d\lambda}{d\nu} \frac{d\nu}{d\mu}.$$

The equalities hold almost everywhere.

**Exercise 12.14.** Suppose a measure  $\nu$  is differentiable with respect to another measure  $\mu$ , and the Radon-Nikodym derivative  $\frac{d\nu}{d\mu} > 0$ . Prove that  $\mu$  is also absolutely continuous with respect to  $\nu$ . Moreover, if  $\mu$  is differentiable with respect to  $\nu$ , then

$$\frac{d\nu}{d\mu} = \left( \frac{d\mu}{d\nu} \right)^{-1}$$

almost everywhere.

## Lebesgue Decomposition

The mutually singular property and the absolutely continuous property are two extreme relations between a signed measure  $\nu$  and a measure  $\mu$ . In general, the relation between the two is something in between.

**Theorem 12.1.9.** Suppose  $\mu$  is a measure and  $\nu$  is  $\sigma$ -finite signed measure. Then there are unique signed measures  $\nu_0$  and  $\nu_1$ , such that

$$\nu = \nu_0 + \nu_1, \quad \nu_0 \perp \mu, \quad \nu_1 \ll \mu.$$

The Lebesgue decomposition may not exist without the  $\sigma$ -finite assumption.

*Proof.* First consider the case  $\nu$  is a measure, i.e.,  $\nu(A) \geq 0$  for all measurable  $A$ . The failure of  $\nu \ll \mu$  is due to the possibility of  $\mu(A) = 0$  and  $\nu(A) > 0$ . We try to show that such failure can be gathered into a measurable subset  $X_0$ . Then within the complement  $X_1 = X - X_0$ , we expect  $\nu \ll \mu$  happens. The restrictions  $\nu_0(A) = \nu(A \cap X_0)$  and  $\nu_1(A) = \nu(A \cap X_1)$  should give the decomposition in the theorem.

The property  $\mu(A) = 0$  and  $\nu(A) > 0$  is the same as  $(\nu - n\mu)(A) > 0$  for any  $n \in \mathbb{N}$ . Therefore by Theorem 12.1.2, we have Hahn decomposition  $X = Y_n^+ \sqcup Y_n^-$  for  $\nu - n\mu$ . Then we take  $X_0 = \cap Y_n^+$  and  $X_1 = X - X_0 = \cup Y_n^-$ .

Assume  $\nu(X)$  is finite. Then for any measurable  $A \subset X_0$ , we have  $0 \leq n\mu(A) < \nu(A) \leq \nu(X)$  for any  $n$ . Since  $\nu(X)$  is a fixed finite number, this implies that  $\mu(A) = 0$ , and further implies that the measure  $\nu_0(A) = \nu(A \cap X_0)$  and  $\mu$  are mutually singular via the decomposition  $X = X_0 \sqcup X_1$ . On the other hand, if  $A \subset X_1$ , then we can write  $A = \sqcup A_n$  with  $A_n \subset Y_n^-$ . By  $A_n \subset Y_n^-$ , we have  $n\mu(A_n) \geq \nu(A_n) \geq 0$ . Therefore

$$\mu(A) = 0 \implies \mu(A_n) = 0 \implies \nu(A_n) = 0 \implies \nu(A) = \sum \nu(A_n) = 0.$$

This shows that the measure  $\nu_1(A) = \nu(A \cap X_1)$  is absolutely continuous with respect to  $\mu$ .

Assume  $\nu$  is  $\sigma$ -finite. Then we have  $X = \sqcup X_i$ , such that  $\nu(X_i) < +\infty$ . We have Lebesgue decomposition  $\nu = \nu_{i0} + \nu_{i1}$  for the restrictions of the measures to measurable subsets in  $X_i$ . Then  $\nu_0(A) = \sum \nu_{i0}(A \cap X_i)$  and  $\nu_1(A) = \sum \nu_{i1}(A \cap X_i)$  give the Lebesgue decomposition on  $X$ .

Assume  $\nu$  is a  $\sigma$ -finite signed measure. Then  $\nu^+$  and  $\nu^-$  are  $\sigma$ -finite measures, and we have Lebesgue decompositions  $\nu^+ = \nu_0^+ + \nu_1^+$  and  $\nu^- = \nu_0^- + \nu_1^-$ . This gives the Lebesgue decomposition  $\nu = (\nu_0^+ - \nu_0^-) + (\nu_1^+ - \nu_1^-)$ . See the second and third parts of Exercise 12.9.

Finally, suppose  $\nu = \nu'_0 + \nu'_1$  is another Lebesgue decomposition. Then we have  $\nu_0 - \nu'_0 \perp \mu$  and  $\nu'_1 - \nu_1 \ll \mu$ . However,  $\nu_0 - \nu'_0 = \nu'_1 - \nu_1$  is a signed measure that is mutually singular to  $\mu$  and is absolutely continuous with respect to  $\mu$ . By the sixth part of Exercise 12.9, this implies that  $\nu_0 - \nu'_0 = \nu'_1 - \nu_1 = 0$  and the uniqueness of the Lebesgue decomposition.

Strictly speaking, the argument for uniqueness requires the values of  $\nu$  to be finite (in order for  $\nu_0 - \nu'_0$  to make sense). The finite case can be extended to the  $\sigma$ -finite case by standard argument.  $\square$

Here is an alternative proof of the Lebesgue decomposition in case both  $\mu$  and  $\nu$  are  $\sigma$ -finite measures. Since  $\mu$  is absolutely continuous with respect to the  $\sigma$ -finite measure  $\lambda = \mu + \nu$ , we have  $\mu(A) = \int_A f d\lambda$  by Radon-Nikodym Theorem. Let  $X_0 = \{x: f(x) = 0\}$  and  $X_1 = \{x: f(x) > 0\}$ . Then  $\nu_0(A) = \nu(A \cap X_0)$  and  $\nu_1(A) = \nu(A \cap X_1)$  give the Lebesgue decomposition.

## 12.2 Lebesgue Differentiation Theorem

In  $\nu(A) = \int_A f d\mu$ , the Radon-Nikodym derivative  $f$  may be considered as the “ratio” between the two measures  $\nu$  and  $\mu$ . If  $f$  is “continuous” at  $a$ , then for a subset  $A$  “close to”  $a$ , we have  $|f(x) - f(a)| < \epsilon$  for small  $\epsilon$  and  $x \in A$ . This implies

$$\left| \frac{\nu(A)}{\mu(A)} - f(a) \right| = \frac{1}{\mu(A)} \left| \int_A (f - f(a)) d\mu \right| \leq \frac{1}{\mu(A)} \int_A |f - f(a)| d\mu \leq \epsilon.$$

This suggests the following formula for computing the Radon-Nikodym derivative

$$f(a) = \lim_{A \rightarrow a} \frac{\nu(A)}{\mu(A)}.$$

There are two problems with the formula. The first is that limit and continuity do not make sense in general measure spaces. The second is that  $f$  is unique only up to a subset of measure zero, and modifying a function on a subset of measure zero may change the continuity dramatically.

If we restrict to Euclidean spaces and the Lebesgue measure  $\mu$ , then the first problem disappears. What we expected actually makes sense and is true.

**Proposition 12.2.1.** *Suppose  $f$  is Lebesgue integrable on  $\mathbb{R}^n$ ,  $\mu$  is the Lebesgue measure on  $\mathbb{R}^n$ , and  $\nu(A) = \int_A f d\mu$ . Then*

$$\lim_{\epsilon \rightarrow 0} \frac{\nu(B(\vec{a}, \epsilon))}{\mu(B(\vec{a}, \epsilon))} = f(\vec{a}) \text{ for almost all } \vec{a} \in \mathbb{R}^n.$$

Suppose  $\nu$  is absolutely continuous with respect to the Lebesgue measure  $\mu$ . If  $\nu(B)$  is finite for any ball  $B$ , then we have  $\nu(A) = \int_A f d\mu$  for a Lebesgue measurable  $f$ , and we may apply the proposition to  $f\chi_B$  for big  $B$  to get the Radon-Nikodym derivative

$$\frac{d\nu}{d\mu}(\vec{a}) = f(\vec{a}) = \lim_{\epsilon \rightarrow 0} \frac{\nu(B(\vec{a}, \epsilon))}{\mu(B(\vec{a}, \epsilon))} \text{ for almost all } \vec{a}.$$

This is exactly what we expected.

Another consequence of the proposition is obtained by taking  $f = \chi_{A \cap B}$  for a Lebesgue measurable subset  $A \subset \mathbb{R}^n$  and big ball  $B$ . We get the *Lebesgue Density Theorem*

$$\lim_{\epsilon \rightarrow 0} \frac{\mu(A \cap B(\vec{a}, \epsilon))}{\mu(B(\vec{a}, \epsilon))} = 1 \text{ for almost all } \vec{a} \in A.$$

This means that any Lebesgue measurable subset  $A$  fills up the small neighborhood of almost every point in  $A$ .

Proposition 12.2.1 is the consequence of a slightly stronger result.

**Theorem 12.2.2 (Lebesgue Differentiation Theorem).** *Suppose  $f$  is Lebesgue integrable on  $\mathbb{R}^n$ , and  $\mu$  is the Lebesgue measure on  $\mathbb{R}^n$ . Then*

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\mu(B(\vec{a}, \epsilon))} \int_{B(\vec{a}, \epsilon)} |f - f(\vec{a})| d\mu = 0 \text{ for almost all } \vec{a} \in \mathbb{R}^n.$$

The theorem implies the proposition because

$$\begin{aligned} \left| \frac{\nu(B(\vec{a}, \epsilon))}{\mu(B(\vec{a}, \epsilon))} - f(\vec{a}) \right| &= \left| \frac{1}{\mu(B(\vec{a}, \epsilon))} \int_{B(\vec{a}, \epsilon)} (f - f(\vec{a})) d\mu \right| \\ &\leq \frac{1}{\mu(B(\vec{a}, \epsilon))} \int_{B(\vec{a}, \epsilon)} |f - f(\vec{a})| d\mu. \end{aligned}$$

**Exercise 12.15.** Suppose  $f$  is Lebesgue integrable on  $\mathbb{R}^n$ . Prove that for almost all  $\vec{a}$ , we have

$$\lim_{A \rightarrow \vec{a}} \frac{1}{\mu(A)} \int_A |f - f(\vec{a})| d\mu = 0.$$

Here the limit at  $\vec{a}$  means that, for any  $\epsilon, \alpha > 0$ , there is  $\delta > 0$ , such that

$$A \subset B(\vec{a}, \epsilon), \mu(A) > \alpha \mu(B(\vec{a}, \epsilon)) \implies \frac{1}{\mu(A)} \int_A |f - f(\vec{a})| d\mu < \epsilon.$$

What does this tell you about the Radon-Nikodym derivative?

**Exercise 12.16.** Suppose  $f(x)$  is a single variable Lebesgue integrable function. Suppose  $F(x) = \int_{x_0}^x f(t) dt$ . Prove that if  $f(x)$  is continuous at  $a$ , then

$$f(a) = \lim_{\substack{x \rightarrow a^- \\ y \rightarrow a^+}} \frac{F(y) - F(x)}{y - x}.$$

Compare the formula with the formula for the Radon-Nikodym derivative.

Now we turn to the proof of the Lebesgue Differentiation Theorem. The argument at the beginning of the section is rigorous on the Euclidean space and proves that the equality in Theorem 12.2.2 holds *everywhere* for a continuous function  $f$ . The idea for the general case is to approximate any Lebesgue measurable function by a continuous function. For functions on  $\mathbb{R}$ , such approximations are given by Proposition 10.5.3. The proposition can be extended to  $\mathbb{R}^n$ .

**Proposition 12.2.3.** *Suppose  $f$  is Lebesgue integrable on  $\mathbb{R}^n$ . Then for any  $\epsilon > 0$ , there is a compactly supported smooth function  $g$ , such that  $\|f - g\|_1 = \int_{\mathbb{R}^n} |f - g| d\mu < \epsilon$ .*

Proposition 10.5.4 can also be extended.

**Proposition 12.2.4.** *Suppose  $f$  is Lebesgue integrable on  $\mathbb{R}$ . If  $\int_{\mathbb{R}^n} f g d\mu = 0$  for any compactly supported smooth function  $g$ , then  $f = 0$  almost everywhere.*

**Exercise 12.17.** Show that for any bounded rectangle  $I$  and  $\epsilon > 0$ , there is a smooth function  $0 \leq g \leq \chi_I$ , such that  $\int_{\mathbb{R}^n} |\chi_I - g| d\mu < \epsilon$ . Then use this to prove Proposition 12.2.3.

**Exercise 12.18.** Prove Proposition 12.2.4.

**Exercise 12.19.** Extend Lusin's Theorem (Theorem 10.5.5) to  $\mathbb{R}^n$ .

Back to the proof of the Lebesgue Differentiation Theorem. Let  $g$  be the

continuous function in Proposition 12.2.3. Let  $h = f - g$ . Then

$$\begin{aligned} & \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |f - f(\vec{a})| d\mu \\ & \leq \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |g - g(\vec{a})| d\mu + \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |h - h(\vec{a})| d\mu \\ & \leq \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |g - g(\vec{a})| d\mu + \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |h| d\mu + |h(\vec{a})|. \end{aligned}$$

Due to the continuity of  $g$ , the  $g$  part can be as small as we want. The  $h$  part should be estimated from  $\|h\|_1 = \|f - g\|_1 < \epsilon$ . The size of the places where  $|h(\vec{a})| \geq \epsilon$  may be estimated by

$$\|h\|_1 \geq \int_{\{|h| \geq \epsilon\}} |h| d\mu \geq \epsilon \mu(\{\vec{a}: |h(\vec{a})| \geq \epsilon\}).$$

Therefore  $\|h\|_1 < \epsilon^2$  implies  $\mu(\{\vec{a}: |h(\vec{a})| \geq \epsilon\}) < \epsilon$ . For the average of  $|h|$  on ball  $B(\vec{a}, \delta)$ , we have the estimation by the *maximal function*  $Mh$  of Hardy and Littlewood

$$\frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |h| d\mu \leq Mh(\vec{a}) = \sup_{r>0} \frac{1}{\mu(B(\vec{a}, r))} \int_{B(\vec{a}, r)} |h| d\mu.$$

**Lemma 12.2.5** (Hardy-Littlewood). *For any Lebesgue integrable  $f$  on  $\mathbb{R}^n$ , we have*

$$\epsilon \mu(\{\vec{a}: Mf(\vec{a}) > \epsilon\}) \leq 3^n \|f\|_1 \text{ for any } \epsilon > 0.$$

*Proof.* The maximal function is actually defined for the measure  $\nu(A) = \int_A |f| d\mu$

$$M\nu(\vec{a}) = \sup_{r>0} \frac{\nu(B(\vec{a}, r))}{\mu(B(\vec{a}, r))}.$$

For  $A_\epsilon = \{\vec{a}: M\nu(\vec{a}) > \epsilon\}$ , the inequality we wish to prove is  $\epsilon \mu(A_\epsilon) \leq 3^n \nu(\mathbb{R}^n)$ .

Implicit in the statement of the proposition is that  $A_\epsilon$  is Lebesgue measurable. This is due to the lower semi-continuity of  $M\nu$  and the obvious extension of Exercise 10.7 to multivariable functions. Note that  $M\nu(\vec{a}) > l$  means  $\nu(B(\vec{a}, r)) > l\mu(B(\vec{a}, r))$  for some  $r$ . Then we expect moving  $\vec{a}$  and  $r$  a little bit will still keep the inequality. Specifically, since  $\mu(B(\vec{a}, r + \delta))$  converges to  $\mu(B(\vec{a}, r))$  as  $\delta \rightarrow 0$ , we have  $\nu(B(\vec{a}, r)) > l\mu(B(\vec{a}, r + \delta))$  for some small  $\delta$ . We may further assume  $r > \delta > 0$ . Then

$$\begin{aligned} \|\vec{x} - \vec{a}\| < \delta & \implies B(\vec{a}, r) \subset B(\vec{x}, r + \delta) \\ & \implies \mu(B(\vec{x}, r + \delta)) \geq \nu(B(\vec{a}, r)) > l\mu(B(\vec{a}, r + \delta)) = \mu(B(\vec{x}, r + \delta)). \end{aligned}$$

The inequality we get on the right implies  $M\nu(\vec{x}) > l$ . This proves the lower semi-continuity of  $M\nu$ .

By Proposition 11.4.3, to prove  $\epsilon\mu(A_\epsilon) \leq 3^n\nu(\mathbb{R}^n)$ , it is sufficient to prove  $\epsilon\mu(K) \leq 3^n\nu(\mathbb{R}^n)$  for any compact subsets  $K \subset A_\epsilon$ . Now  $\vec{a} \in A_\epsilon$  means that  $\nu(B(\vec{a}, r_{\vec{a}})) > \epsilon\mu(B(\vec{a}, r_{\vec{a}}))$  for a radius  $r_{\vec{a}}$ . Then  $K$  is covered by the collection of open balls  $B(\vec{a}, r_{\vec{a}})$ ,  $\vec{a} \in K$ . Since  $K$  is compact, it is covered by finitely many such balls

$$K \subset B_1 \cup \cdots \cup B_k, \quad B_i = B(\vec{a}_i, r_i), \quad r_i = r_{\vec{a}_i}.$$

Then we have *Vitali's Covering Lemma*: There are disjoint  $B_{i_1}, \dots, B_{i_l}$ , such that

$$B_1 \cup \cdots \cup B_k \subset 3B_{i_1} \cup \cdots \cup 3B_{i_l}, \quad 3B_i = B(\vec{a}_i, 3r_i).$$

To find these disjoint balls, we first assume  $r_1 \geq \cdots \geq r_k$ , without loss of generality. We first choose  $B_{i_1} = B_1$ . Then after choosing  $B_{i_{j-1}}$ , we choose  $B_{i_j}$  to be the first  $B_i$  disjoint from  $B_{i_1}, \dots, B_{i_{j-1}}$ . The choice implies that any  $B_i$  intersects some  $B_{i_j}$  with  $i_j \leq i$ . Since the radius  $r_{i_j}$  of  $B_{i_j}$  is bigger than the radius  $r_i$  of  $B_i$ , and the two balls intersect, enlarging the bigger ball  $B_{i_j}$  three times will swallow the smaller ball  $B_i$ . In other words, we have  $B_i \subset 3B_{i_1}$ . Then we conclude

$$\epsilon\mu(K) \leq \epsilon \sum_j \mu(3B_{i_j}) < 3^n \sum_j \nu(B_{i_j}) = 3^n \nu(\sqcup_j B_{i_j}) \leq 3^n \nu(\mathbb{R}^n). \quad \square$$

*Proof of Theorem 12.2.2.* For any  $1 > \epsilon > 0$ , by Proposition 12.2.3, there is a compactly supported continuous function  $g$ , such that  $\|f - g\|_1 < \epsilon^2$ . Then  $g$  is uniformly continuous. This means that there is  $\delta_\epsilon > 0$ , such that

$$\|\vec{x} - \vec{y}\| < \delta_\epsilon \implies |g(\vec{x}) - g(\vec{y})| < \epsilon.$$

Then

$$0 < \delta < \delta_\epsilon \implies \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |g - g(\vec{a})| d\mu \leq \epsilon.$$

The function  $h = f - g$  satisfies  $\|h\|_1 \leq \epsilon^2$ . By Lemma 12.2.5, we have

$$\mu(\{\vec{a}: |Mh(\vec{a})| > \epsilon\}) \leq \frac{3^n}{\epsilon} \|h\|_1 \leq 3^n \epsilon.$$

We also have

$$\mu(\{\vec{a}: |h(\vec{a})| \geq \epsilon\}) \leq \frac{1}{\epsilon} \|h\|_1 \leq \epsilon.$$

For  $0 < \delta < \delta_\epsilon$ , we then have

$$\begin{aligned} & \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |f - f(\vec{a})| d\mu \\ & \leq \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |g - g(\vec{a})| d\mu + \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |h| d\mu + |h(\vec{a})| \\ & \leq \epsilon + |Mh(\vec{a})| + |h(\vec{a})|. \end{aligned}$$



If the left side is  $\geq 3\epsilon$ , then either  $|Mh(\vec{a})| > \epsilon$  or  $|h(\vec{a})| \geq \epsilon$ . This means

$$X_\epsilon = \left\{ \vec{a}: \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |f - f(\vec{a})| d\mu \geq 3\epsilon \text{ for some } 0 < \delta < \delta_\epsilon \right\} \\ \subset \{ \vec{a}: |Mh(\vec{a})| > \epsilon \} \cup \{ \vec{a}: |h(\vec{a})| \geq \epsilon \}.$$

This implies  $\mu(X_\epsilon) \leq (1 + 3^n)\epsilon$ .

Let  $\epsilon_k = \frac{1}{2^k}$  and  $\delta_k = \delta_{\epsilon_k}$ . Then  $Y = \cap_{m=1}^\infty \cup_{k>m} X_{\epsilon_k}$  has zero measure. For  $\vec{a} \notin Y$ , there is  $m$ , such that

$$0 < \delta < \delta_k, k > m \implies \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |f - f(\vec{a})| d\mu < 3\epsilon_k.$$

Since  $\epsilon_k$  can be arbitrarily small for  $k > m$ , the above further implies

$$\lim_{\delta \rightarrow 0^+} \frac{1}{\mu(B(\vec{a}, \delta))} \int_{B(\vec{a}, \delta)} |f - f(\vec{a})| d\mu = 0. \quad \square$$

## 12.3 Differentiation on $\mathbb{R}$ : Fundamental Theorem

What function is the integration of another function:  $f(x) = f(a) + \int_a^x g(t) dt$ ? The Fundamental Theorem of Riemann Integral tells us that, if  $g$  is Riemann integrable, then  $f$  is continuous in general, and is differentiable and satisfies  $f' = g$  wherever  $g$  is continuous. Moreover, if  $f$  is differentiable and the derivative  $f'$  is Riemann integrable, then the equality holds for  $g = f'$ .

For Lebesgue integrable  $g$ , we actually have a signed measure  $\nu(A) = \int_A g d\mu$ , where  $\mu$  is the usual Lebesgue measure. Moreover, the integral function is the *distribution function*

$$f(x) = \int_{-\infty}^x g d\mu = \nu(-\infty, x),$$

and  $\nu$  should be the signed version of the Lebesgue-Stieltjes measure induced by  $f$ . Therefore the Fundamental Theorem of Lebesgue Integral can be put into the bigger context of the Radon-Nikodym derivative of absolutely continuous Lebesgue-Stieltjes measure.

### Signed Lebesgue-Stieltjes Measure

We extend the Lebesgue-Stieltjes measure developed in Section 11.2 to signed case. A bounded variation function  $f$  has the positive and negative variation functions  $v^+$  and  $v^-$ . Then the *signed Lebesgue-Stieltjes measure* induced by  $f$  is  $\mu_f = \mu_{v^+} - \mu_{v^-}$ . This means ( $a$  and  $b$  are decorated with  $\pm$ )

$$\begin{aligned} \mu_f \langle a, b \rangle &= \mu_{v^+} \langle a, b \rangle - \mu_{v^-} \langle a, b \rangle \\ &= (v^+(b) - v^+(a)) - (v^-(b) - v^-(a)) \\ &= f(b) - f(a). \end{aligned}$$

Implicit in the definition is that a subset is Lebesgue-Stieltjes measurable with respect to  $f$  if and only if it is Lebesgue-Stieltjes measurable with respect to  $v^+$  and  $v^-$ . For our purpose, it is sufficient to restrict ourselves to Borel sets, and think of the general measurability with respect to  $f$  as given by the completion with respect to  $\mu_v = \mu_{v^+} + \mu_{v^-}$  (see Exercise 11.20).

**Example 12.3.1.** Consider

$$f(x) = \begin{cases} 0, & \text{if } x \neq a, \\ 1, & \text{if } x = a, \end{cases} \quad v^+(x) = \begin{cases} 0, & \text{if } x < a, \\ 1, & \text{if } x \geq a, \end{cases} \quad v^-(x) = \begin{cases} 0, & \text{if } x \leq a, \\ 1, & \text{if } x > a. \end{cases}$$

Both  $v^+$  and  $v^-$  induce the Dirac measure concentrated at  $a$ , introduced in Example 9.4.5. Therefore the Lebesgue-Stieltjes measure with respect to  $f$  is always zero. In particular,  $\mu_f = \mu_{v^+} - \mu_{v^-}$  is not the Jordan decomposition.

The reason for this peculiar example is that the Lebesgue-Stieltjes measure is really defined by the left or right limits instead of the value of the function itself. See Exercises 11.22 and 12.22. Therefore  $f(x)$ ,  $f(x^+)$  and  $f(x^-)$  should induce the same signed Lebesgue-Stieltjes measure.

**Exercise 12.20.** Suppose  $f$  has bounded variation, with positive variation  $v^+$  and negative variation  $v^-$ . Suppose  $f = u^+ - u^-$  for increasing functions  $u^+, u^-$ .

1. Prove that if  $A$  is Lebesgue-Stieltjes measurable with respect to  $u^+$ , then  $A$  is Lebesgue-Stieltjes measurable with respect to  $v^+$ , and  $\mu_{v^+}(A) \leq \mu_{u^+}(A)$ .
2. Prove that if  $A$  is Lebesgue-Stieltjes measurable with respect to  $u^+$  and  $u^-$ , then  $\mu_f(A) = \mu_{u^+}(A) - \mu_{u^-}(A)$ .

**Exercise 12.21.** Suppose  $f$  has bounded variation, with variation function  $v$ . Prove that  $A$  is Lebesgue-Stieltjes measurable with respect to  $f$  if and only if it is  $A$  is Lebesgue-Stieltjes measurable with respect to  $v$ . However, it is not necessarily true that  $|\mu_f| = \mu_v$ .

**Exercise 12.22.** Suppose  $f$  is a bounded variation function. Prove that  $f(x^+)$  also has bounded variation and  $\mu_{f(x)} = \mu_{f(x^+)}$ .

**Exercise 12.23.** Suppose  $f$  has bounded variation, with positive variation  $v^+$  and negative variation  $v^-$ . Prove that if  $f(x)$  is between  $f(x^-)$  and  $f(x^+)$  for all  $x$ , then  $\mu_f = \mu_{v^+} - \mu_{v^-}$  is the Jordan decomposition.

With the Lebesgue-Stieltjes measure extended to bounded variation functions, we can compare the Fundamental Theorem of Calculus with the Radon-Nikodym Theorem with respect to the usual Lebesgue measure.

**Proposition 12.3.1.** *Suppose  $f$  has bounded variation and is continuous. Suppose  $g$  is a Lebesgue integrable function. Then  $f(x) = f(a) + \int_a^x g d\mu$  for all  $x$  if and only if  $\mu_f(A) = \int_A g d\mu$  for all Borel set  $A$ .*

The proposition can be applied to whole  $\mathbb{R}$  or an interval. Proposition 12.3.1 shows the necessity of the continuity assumption. In fact, the proposition holds as long as  $f(x)$  is assumed between  $f(x^-)$  and  $f(x^+)$  for all  $x$ . However, by Exercise 10.31, the function  $f$  will always be continuous at the end.

*Proof.* If  $\mu_f(A) = \int_A g d\mu$  and  $f$  is continuous at  $a$  and  $x$ , then applying the equality to  $A = \langle a, x \rangle$  gives  $f(x) - f(a) = \int_a^x g d\mu$ .

Conversely, suppose  $f(x) = f(a) + \int_a^x g d\mu$ , we want to prove that  $\mu_f$  is the same as the signed measure  $\nu(A) = \int_A g d\mu$ . We have  $\mu_f(x, y) = f(y) - f(x) = \int_{(x, y)} g d\mu = \nu(x, y)$ . By taking countable disjoint union of open intervals, we get  $\mu_f(U) = \nu(U)$  for any open subset  $U$ .

Assume  $g \geq 0$  and  $A$  is a bounded Borel set. For any  $\epsilon > 0$ , by Proposition 11.2.1, there is open  $U \supset A$ , such that  $\mu_f(U - A) < \epsilon$ . Moreover, by Proposition 12.1.5 (or Exercise 10.3.9), there is open  $V \supset A$ , such that  $\nu(V - A) < \epsilon$ . Then

$$0 \leq \mu_f(U \cap V) - \mu_f(A) = \mu_f(U \cap V - A) \leq \mu_f(U - A) < \epsilon,$$

and

$$0 \leq \nu(U \cap V) - \nu(A) = \nu(U \cap V - A) \leq \nu(U - A) < \epsilon.$$

We have  $\mu_f(U \cap V) = \nu(U \cap V)$  for the open subset  $U \cap V$ . Therefore  $|\mu_f(A) - \nu(A)| < \epsilon$ . Since  $\epsilon$  is arbitrary, we get  $\mu_f(A) = \nu(A)$ .

The equality  $\mu_f(A) = \nu(A)$  for unbounded  $A$  can be obtained by adding the equalities  $\mu_f(A \cap (i, i+1]) = \nu(A \cap (i, i+1])$  together. For general  $g$ , we prove the equalities for  $g^+ = \max\{g, 0\}$  and  $g^- = -\min\{g, 0\}$  and then subtract the two.  $\square$

## Absolutely Continuous Function

By Radon-Nikodym Theorem (Theorem 12.1.6) and Proposition 12.3.1, a continuous bounded variation function is the integral of a Lebesgue integrable function if and only if the corresponding Lebesgue-Stieltjes measure is absolutely continuous with respect to the usual Lebesgue measure  $\mu$ . We wish to know what the absolute continuity means in terms of the original function  $f$ .

**Proposition 12.3.2.** *Suppose  $f$  is a continuous bounded variation function. Then the Lebesgue-Stieltjes measure  $\mu_f$  is absolutely continuous with respect to the Lebesgue measure if and only if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that for disjoint intervals  $(x_i, y_i)$ , we have*

$$|x_1 - y_1| + \cdots + |x_n - y_n| < \delta \implies |f(x_1) - f(y_1)| + \cdots + |f(x_n) - f(y_n)| < \epsilon.$$

Naturally, a function satisfying the property in the proposition is called an *absolutely continuous* function. The concept is defined over any (bounded or unbounded) interval.

The proof will show that absolutely continuous functions have bounded variation. Moreover, the variation  $v$ , positive variation  $v^+$  and negative variation  $v^-$  are also absolutely continuous. The proof will also show that the implication is true for infinitely many disjoint intervals

$$\sum_{i=1}^{\infty} |x_i - y_i| < \delta \implies \sum_{i=1}^{\infty} |f(x_i) - f(y_i)| < \epsilon.$$

It is critical to allow  $n$  to be arbitrarily big. If we only allow  $n = 1$ , then the definition becomes the uniform continuity. In fact, if we impose a bound on  $n$ , the definition is still equivalent to the uniform continuity. We will see in Example 12.3.2 that the Cantor function in Example 11.4.5 is continuous (and therefore uniformly continuous on  $[0, 1]$ ) but is not absolutely continuous.

It is also critical to require the intervals to be disjoint. Example 12.3.3 will show that  $\sqrt{x}$  is absolutely continuous but does not allow the intervals to overlap. In fact, Exercise 12.26 will show that allowing overlapping means exactly that  $f$  is a Lipschitz function.

In summary, we have the following strict relations

$$\begin{aligned} \text{Lipschitz} &> \text{absolutely continuous} \\ &> \text{uniformly continuous and bounded variation.} \end{aligned}$$

*Proof.* Suppose  $\mu_f$  is absolutely continuous. Then by Proposition 12.1.5, for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$\mu(A) < \delta \implies |\mu_f(A)| < \epsilon.$$

For the case that  $A$  is an open subset  $U = \sqcup(x_i, y_i)$ , we get

$$\mu(U) = \sum (y_i - x_i) < \delta \implies |\mu_f(U)| = \left| \sum (f(y_i) - f(x_i)) \right| < \epsilon.$$

The left side is  $\sum |x_i - y_i|$ . But the right side is not  $\sum |f(x_i) - f(y_i)|$ . To move the absolute value of the right side to the inside of  $\sum$ , we write

$$U = U^+ \sqcup U^-, \quad U^+ = \sqcup_{f(y_i) \geq f(x_i)} (x_i, y_i), \quad U^- = \sqcup_{f(y_i) < f(x_i)} (x_i, y_i).$$

Then  $\mu(U) = \sum |x_i - y_i| < \delta$  implies  $\mu(U^+) < \delta$  and  $\mu(U^-) < \delta$ , so that

$$\begin{aligned} |\mu_f(U^+)| &= \left| \sum_{f(y_i) \geq f(x_i)} (f(y_i) - f(x_i)) \right| = \sum_{f(y_i) \geq f(x_i)} |f(x_i) - f(y_i)| < \epsilon, \\ |\mu_f(U^-)| &= \left| \sum_{f(y_i) < f(x_i)} (f(y_i) - f(x_i)) \right| = \sum_{f(y_i) < f(x_i)} |f(x_i) - f(y_i)| < \epsilon. \end{aligned}$$

Then we get

$$\sum |f(x_i) - f(y_i)| = \sum_{f(y_i) \geq f(x_i)} |f(x_i) - f(y_i)| + \sum_{f(y_i) < f(x_i)} |f(x_i) - f(y_i)| < 2\epsilon.$$

This proves the necessity direction of the proposition (up to substituting  $\epsilon$  by  $2\epsilon$ ).

For the sufficiency direction, we consider the variation  $v$ , positive variation  $v^+$  and negative variation  $v^-$  of  $f$ . We first show that the  $\epsilon$ - $\delta$  statement for  $f$  implies the same  $\epsilon$ - $\delta$  statement for the three variation functions.

Let  $U = \sqcup(x_i, y_i)$  be an open subset with  $\mu(U) = \sum(y_i - x_i) < \delta$ . Suppose each interval  $(x_i, y_i)$  has a partition

$$P_i: x_i = z_{i0} < z_{i1} < \cdots < z_{in_i} = y_i.$$

Then all intervals  $(z_{i(j-1)}, z_{ij})$  are disjoint and have total length

$$\sum_i \sum_j |z_{i(j-1)} - z_{ij}| = \mu(U) < \delta.$$

By the  $\epsilon$ - $\delta$  statement for  $f$ , we have

$$\sum_i V_{P_i}(f) = \sum_i \sum_j |f(z_{i(j-1)}) - f(z_{ij})| < \epsilon.$$

By taking  $P_i$  to be more and more refined, we get

$$\sum V_{(x_i, y_i)}(f) = \sum |v(x_i) - v(y_i)| \leq \epsilon.$$

This is the  $\epsilon$ - $\delta$  statement for  $v$ . By  $|v^+(x) - v^+(y)| \leq |v(x) - v(y)|$  and  $|v^-(x) - v^-(y)| \leq |v(x) - v(y)|$ , this further implies the  $\epsilon$ - $\delta$  statement for  $v^+$  and  $v^-$ .

We remark that in the special case  $U = (x, y)$ , the  $\epsilon$ - $\delta$  statement for  $v$  says that the variation of  $f$  on any interval of length  $\delta$  is bounded by  $\epsilon$ . Since any bounded interval can be divided into finitely many intervals of length  $\delta$ , we find that  $f$  has bounded variation on any bounded interval.

The  $\epsilon$ - $\delta$  statement for  $v^+$  and  $v^-$  also implies that  $\mu_f$  is absolutely continuous with respect to  $\mu$ . Suppose  $\mu(A) = 0$ . Then for  $\delta$  in the  $\epsilon$ - $\delta$  statement for  $v$ , we find

$$A \subset U = \sqcup(x_i, y_i), \quad \mu(U) = \sum |x_i - y_i| < \delta.$$

Applying the  $\epsilon$ - $\delta$  statement for  $v^+$  to  $U$  (strictly speaking, first to finite unions in  $U$  and then taking limit), we get

$$0 \leq \mu_{v^+}(A) \leq \mu_{v^+}(U) = \sum |v^+(x_i) - v^+(y_i)| < \epsilon.$$

Since  $\epsilon$  is arbitrary, we get  $\mu_{v^+}(A) = 0$ . By the same reason, we have  $\mu_{v^-}(A) = 0$ . Therefore  $\mu_f(A) = 0$ .  $\square$

**Example 12.3.2.** The Cantor function  $\kappa$  in Example 11.4.5 is continuous. We show that it is not absolutely continuous. By Proposition 12.3.2, this means that the Lebesgue-Stieltjes measure  $\mu_\kappa$  is not absolutely continuous with respect to the usual Lebesgue measure. In fact, Example 12.1.2 shows that they are mutually singular.

We have  $K \subset K_n$ , where  $K_n$  is a union of  $2^n$  closed intervals  $[x_i, y_i]$  of length  $\frac{1}{3^n}$  in Example 9.2.4. We have  $\mu(K_n) = \sum |x_i - y_i| = \frac{2^n}{3^n}$  converging to 0. On the other hand, since  $\kappa$  is constant on the intervals  $[0, 1] - K_n$ , we see that  $\sum (f(y_i) - f(x_i)) = 1$ . Therefore the Cantor function is not absolutely continuous.

**Example 12.3.3.** Consider the function  $\sqrt{x}$  on  $[0, +\infty)$ . For any  $1 > \epsilon > 0$ , we have

$$y > x \geq \epsilon^2 \implies |\sqrt{x} - \sqrt{y}| \leq \frac{1}{2\sqrt{c}}|x - y| \leq \frac{1}{2\epsilon}|x - y|, \text{ for some } c \in (x, y).$$

Now consider a collection of disjoint open intervals  $(x_i, y_i)$  satisfying  $\sum |x_i - y_i| < \epsilon^2$ . If  $\epsilon^2$  belongs to an interval  $(x_j, y_j)$ , then we may break the interval into  $(x_j, \epsilon^2)$  and  $(\epsilon^2, y_j)$ , and this does not affect the discussion. Therefore we assume that any  $(x_i, y_i)$  is contained in either  $(0, \epsilon^2)$  or  $(\epsilon^2, +\infty)$  and get

$$\begin{aligned} \sum |\sqrt{x_i} - \sqrt{y_i}| &= \sum_{(x_i, y_i) \subset (0, \epsilon^2)} |\sqrt{x_i} - \sqrt{y_i}| + \sum_{(x_i, y_i) \subset (\epsilon^2, +\infty)} |\sqrt{x_i} - \sqrt{y_i}| \\ &\leq \sum_{(x_i, y_i) \subset (0, \epsilon^2)} (\sqrt{y_i} - \sqrt{x_i}) + \sum_{(x_i, y_i) \subset (\epsilon^2, +\infty)} \frac{1}{2\epsilon}(y_i - x_i) \\ &\leq \epsilon + \frac{1}{2\epsilon} \sum (y_i - x_i) < \frac{3}{2}\epsilon, \end{aligned}$$

where the second inequality makes use of the disjoint property among the intervals. This shows that  $\sqrt{x}$  is absolutely continuous.

The disjoint condition cannot be removed from the definition. For  $x_i = \frac{1}{n^2}$ ,  $y_i = \frac{4}{n^2}$ ,  $i = 1, 2, \dots, n$ , we have

$$\sum |x_i - y_i| = n \frac{3}{n^2} \rightarrow 0, \quad \sum |\sqrt{x_i} - \sqrt{y_i}| = n \frac{1}{n} = 1.$$

**Exercise 12.24.** Prove that the sum of two absolutely continuous functions is absolutely continuous. Prove that the product of two bounded absolutely continuous functions is absolutely continuous.

**Exercise 12.25.** Prove that if  $f$  is absolutely continuous, and  $g$  is strictly monotone and absolutely continuous, then the composition  $f \circ g$  is absolutely continuous.

**Exercise 12.26.** Prove that a function  $f$  is Lipschitz if and only if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that

$$|x_1 - y_1| + \dots + |x_n - y_n| < \delta \implies |f(x_1) - f(y_1)| + \dots + |f(x_n) - f(y_n)| < \epsilon.$$

Note that this drops the disjoint condition from the definition of the absolute continuity. In particular, Lipschitz functions are absolutely continuous.

**Exercise 12.27.** Prove that if  $f$  is Lipschitz and  $g$  is absolutely continuous, then the composition  $f \circ g$  is absolutely continuous.

**Exercise 12.28.** Show that the composition of two absolutely continuous functions may not be absolutely continuous.

## Differentiability of Monotone Function

So far we have addressed the problem of expressing  $f$  as the integral of some  $g$ . From the fundamental theorem of Riemann integral calculus, we expect  $g$  to be the derivative  $f'$ . Here we establish the existence of  $f'$ .

The proof of Proposition 12.3.2 shows that any absolutely continuous function must have bounded variation. Since any bounded variation function is the difference of two increasing functions, the following result implies the existence of  $f'$  almost everywhere, which is enough for taking the integral.

**Theorem 12.3.3 (Lebesgue).** *An increasing function  $f$  is differentiable almost everywhere. Moreover,  $f'$  is Lebesgue integrable on any bounded interval and satisfies*

$$\int_a^b f' d\mu \leq f(b) - f(a).$$

*Proof.* Introduce the *Dini derivatives*

$$\begin{aligned} D^+ f(x) &= \overline{\lim}_{y \rightarrow x^+} \frac{f(y) - f(x)}{y - x}, & D_+ f(x) &= \underline{\lim}_{y \rightarrow x^+} \frac{f(y) - f(x)}{y - x}, \\ D^- f(x) &= \overline{\lim}_{y \rightarrow x^-} \frac{f(y) - f(x)}{y - x}, & D_- f(x) &= \underline{\lim}_{y \rightarrow x^-} \frac{f(y) - f(x)}{y - x}. \end{aligned}$$

The function is differentiable at  $x$  if and only if  $D^+ f(x) \leq D_- f(x)$  and  $D_+ f(x) \geq D^- f(x)$ . We will prove that the places where  $D^+ f(x) > D_- f(x)$  has measure zero. Similar argument shows that the places where  $D_+ f(x) < D^- f(x)$  also has measure zero.

We have

$$\{x: D^+ f(x) > D_- f(x)\} = \cup_{p, q \in \mathbb{Q}, p > q} \{x: D^+ f(x) > p > q > D_- f(x)\}.$$

To show the left side has measure zero, it suffices to show that, for each pair  $p > q$ , the subset

$$E = \{x: D^+ f(x) > p > q > D_- f(x)\}$$

has measure zero.

For any  $\epsilon > 0$ , we have  $\mu(U) < \mu^*(E) + \epsilon$  for some open  $U \supset E$ . For any  $x \in E$ , by  $D_- f(x) < q$ , there are arbitrarily small closed intervals  $[y, x] \subset U$ , such that  $f(x) - f(y) < q(x - y)$ . Then by Vitali covering lemma (Lemma 12.3.4), we have disjoint  $[y_1, x_1], \dots, [y_n, x_n]$ , such that

$$f(x_i) - f(y_i) < q(x_i - y_i), \quad \mu^*(E - \sqcup_{i=1}^n [y_i, x_i]) < \epsilon.$$

For any  $u \in E' = E \cap (\sqcup_{i=1}^n (y_i, x_i))$ , by  $D^+f(u) > p$ , there are arbitrarily small closed intervals  $[u, v] \subset \sqcup_{i=1}^n (y_i, x_i)$ , such that  $f(v) - f(u) > p(v - u)$ . Applying Vitali covering lemma again, we have disjoint  $[u_1, v_1], \dots, [u_m, v_m]$ , such that

$$f(v_j) - f(u_j) > p(v_j - u_j), \quad \mu^*(E' - \sqcup_{j=1}^m [u_j, v_j]) < \epsilon.$$

Now we carry out estimations. We have

$$\begin{aligned} q \sum (x_i - y_i) &\geq \sum (f(x_i) - f(y_i)) && (f(x_i) - f(y_i) < q(x_i - y_i)) \\ &\geq \sum (f(v_j) - f(u_j)) && (\sqcup [u_j, v_j] \subset \sqcup (y_i, x_i) \text{ and } f \text{ increasing}) \\ &\geq p \sum (v_j - u_j). && (f(v_j) - f(u_j) > p(v_j - u_j)) \end{aligned}$$

The left side is the measure of  $\sqcup [y_i, x_i] \subset U$

$$\sum (x_i - y_i) = \mu(\sqcup [y_i, x_i]) \leq \mu(U) < \mu^*(E) + \epsilon.$$

The right side is the measure of  $\sqcup [u_j, v_j]$ . By

$$E = E' \cup (E - \sqcup_{i=1}^n (y_i, x_i)) \subset (\sqcup [u_j, v_j]) \cup (E' - \sqcup [u_j, v_j]) \cup (E - \sqcup_{i=1}^n (y_i, x_i)),$$

we have (note that changing  $(y_i, x_i)$  to  $[y_i, x_i]$  does not change outer measure)

$$\begin{aligned} \sum (v_j - u_j) &= \mu(\sqcup [u_j, v_j]) \geq \mu^*(E) - \mu^*(E' - \sqcup [u_j, v_j]) - \mu^*(E - \sqcup_{i=1}^n (y_i, x_i)) \\ &> \mu^*(E) - 2\epsilon. \end{aligned}$$

Combining the estimations, we get

$$q(\mu^*(E) + \epsilon) \geq p(\mu^*(E) - 2\epsilon).$$

Since  $p > q$  and  $\epsilon$  is arbitrary, we conclude that  $\mu^*(E) = 0$ .

As explained earlier, we have proved that  $f$  is differentiable almost everywhere. Next we study the integral of the derivative function.

Let  $\epsilon_n > 0$  converge to 0. Then we have

$$f'(x) = \lim_{n \rightarrow \infty} \frac{f(x + \epsilon_n) - f(x)}{\epsilon_n}$$

almost everywhere. Here we extend  $f(x)$  to the whole  $\mathbb{R}$  by  $f = f(a)$  on  $(-\infty, a]$  and  $f = f(b)$  on  $[b, +\infty)$ . The function  $f'$  is Lebesgue measurable because each  $\frac{f(x + \epsilon_n) - f(x)}{\epsilon_n}$  is Lebesgue measurable. Moreover, we have  $\frac{f(x + \epsilon_n) - f(x)}{\epsilon_n} \geq 0$  because  $f$  is increasing. Then we may apply Fatou's Lemma (Theorem 10.4.3) to get

$$\begin{aligned} \int_a^b f' d\mu &= \int_a^b \lim_{n \rightarrow \infty} \frac{f(x + \epsilon_n) - f(x)}{\epsilon_n} d\mu \leq \liminf_{n \rightarrow \infty} \int_a^b \frac{f(x + \epsilon_n) - f(x)}{\epsilon_n} d\mu \\ &= \liminf_{n \rightarrow \infty} \frac{1}{\epsilon_n} \left( \int_{a+\epsilon_n}^{b+\epsilon_n} f d\mu - \int_a^b f d\mu \right) = \liminf_{n \rightarrow \infty} \frac{1}{\epsilon_n} \left( \int_b^{b+\epsilon_n} f d\mu - \int_a^{a+\epsilon_n} f d\mu \right) \\ &\leq f(b) - f(a). \end{aligned}$$

The last inequality is due to  $f = f(b)$  on  $[b, b + \epsilon_n]$  and  $f \geq f(a)$  on  $[a, a + \epsilon_n]$ .  $\square$



**Lemma 12.3.4** (Vitali Covering Theorem). *Suppose  $A \subset \mathbb{R}$  has finite Lebesgue outer measure. Suppose  $\mathcal{V}$  is a collection of intervals, such that for each  $x \in A$  and  $\delta > 0$ , there is  $I \in \mathcal{V}$  satisfying  $x \in I$  and  $\mu(I) < \delta$ . Then for any  $\epsilon > 0$ , there are disjoint  $I_1, \dots, I_n \in \mathcal{V}$ , such that  $\mu^*(A - \sqcup_{i=1}^n I_i) < \epsilon$ .*

*Proof.* By adding end points to intervals, we may assume that all intervals in  $\mathcal{V}$  are closed. Moreover, We may assume all intervals are contained in an open subset  $U$  of finite measure.

We inductively choose intervals  $I_n$  in the “greedy way”. Suppose disjoint intervals  $I_1, \dots, I_{n-1} \in \mathcal{V}$  have been chosen. Let

$$a_n = \sup\{\mu(I) : I \in \mathcal{V}, I \cap I_i = \emptyset \text{ for } i = 1, 2, \dots, n-1\}.$$

Then choose an  $I_n$  in the collection on the right, such that  $\mu(I_n) > \frac{1}{2}a_n$ . By  $\sum \mu(I_n) = \mu(\sqcup I_n) \leq \mu(U) < +\infty$ , the series  $\sum \mu(I_n)$  converges, which implies  $\lim a_n = 0$ . For any  $\epsilon > 0$ , we can then find  $n$ , such that  $\sum_{i>n} \mu(I_i) < \epsilon$ . We wish to estimate the outer measure of  $B = A - \sqcup_{i=1}^n I_i$ .

Since intervals in  $\mathcal{V}$  are closed, for any  $x \in B$ , there is  $\delta > 0$ , such that  $(x - \delta, x + \delta)$  is disjoint from any of  $I_1, \dots, I_n$ . Then by the assumption, we have  $x \in I$  for some  $I \in \mathcal{V}$  satisfying  $\mu(I) < \delta$ . Thus  $I \subset (x - \delta, x + \delta)$  and is disjoint from any of  $I_1, \dots, I_n$ .

By  $\lim a_n = 0$ , we have  $\mu(I) > a_i$  for some  $i$ . By the definition of  $a_i$ , we have  $I \cap I_j \neq \emptyset$  for some  $j \leq i$ . Let  $j$  be the smallest  $j$  satisfying  $I \cap I_j \neq \emptyset$ . Then  $j > n$  (since  $I$  is disjoint from  $I_1, \dots, I_n$ ) and  $\mu(I) \leq a_j$  (since  $I$  is disjoint from  $I_1, \dots, I_{j-1}$ ). By  $I \cap I_j \neq \emptyset$ ,  $\mu(I) \leq a_j$  and  $\mu(I_j) > \frac{1}{2}a_j$ , we have  $I \subset 5I_j$ , where  $5I_j$  is the interval with the same center as  $I_j$  but five times the length. Therefore  $x \in I \subset 5I_j$ , with  $j > n$ . This implies  $B \subset \cup_{j>n} 5I_j$ , and we have

$$\mu^*(B) \leq \mu(\cup_{j>n} 5I_j) = \sum_{j>n} \mu(5I_j) = 5 \sum_{j>n} \mu(I_j) < 5\epsilon. \quad \square$$

**Example 12.3.4.** Since the Cantor function  $\kappa$  is constant on any interval in the complement the Cantor set  $K$ , we get  $\kappa' = 0$  away from  $K$ . Since  $\mu(K) = 0$ , we see that  $\kappa' = 0$  almost everywhere, yet  $\kappa$  is not a constant function. We have  $\int_0^1 f' dx = 0$ ,  $f(1) - f(0) = 1$ , and the inequality in Theorem 12.3.3 becomes strict.

**Exercise 12.29.** Suppose  $f$  is an increasing function on  $[a, b]$  satisfying  $f(b) - f(a) = \int_a^b f' d\mu$ . Prove that  $f(y) - f(x) = \int_x^y f' d\mu$  for any interval  $[x, y]$  inside  $[a, b]$ .

## Fundamental Theorem of Lebesgue Integral

**Theorem 12.3.5** (Fundamental Theorem). *A function on an interval is the integral of a Lebesgue integrable function if and only if it is bounded and absolutely continuous. Moreover, the function is differentiable almost everywhere, and is the integral of its derivative function.*

*Proof.* Radon-Nikodym Theorem (Theorem 12.1.6), Proposition 12.3.1 and Proposition 12.3.2 together tell us that  $f(x) = f(a) + \int_a^x g d\mu$  for some Lebesgue integrable  $g$  if and only if  $f$  is bounded and absolutely continuous.

It remains to show that  $g = f'$  almost everywhere. Here by the proof of Proposition 12.3.2, an absolutely continuous  $f$  must have bounded variation on any bounded interval. Then Theorem 12.3.3 further implies that  $f'$  exists almost everywhere.

We may apply Proposition 12.2.1 to “compute”  $g$ . For almost all  $x$ , the equality in Proposition 12.2.1 holds at  $x$ , and  $f'(x)$  exists. At such an  $x$ , we have

$$\begin{aligned} g(x) &= \lim_{\epsilon \rightarrow 0} \frac{\int_{(x-\epsilon, x+\epsilon)} g d\mu}{\mu(x-\epsilon, x+\epsilon)} = \lim_{\epsilon \rightarrow 0} \frac{f(x+\epsilon) - f(x-\epsilon)}{2\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{2} \left( \frac{f(x+\epsilon) - f(x)}{\epsilon} + \frac{f(x) - f(x-\epsilon)}{\epsilon} \right) \\ &= \frac{1}{2} (f'(x) + f'(x)) = f'(x). \end{aligned}$$

This proves that  $g = f'$  almost everywhere.

We may also prove  $g = f'$  by adopting the last part of the proof of Theorem 12.3.3. This direct proof does not use Proposition 12.2.1.

First assume  $g$  is bounded. We have  $f'(x) = \lim_{n \rightarrow \infty} \frac{f(x+\epsilon_n) - f(x)}{\epsilon_n}$  for a positive sequence  $\epsilon_n \rightarrow 0$ . If  $|g| \leq M$ , then

$$\left| \frac{f(x+\epsilon_n) - f(x)}{\epsilon_n} \right| = \left| \frac{1}{\epsilon_n} \int_x^{x+\epsilon_n} g d\mu \right| < M.$$

Therefore we may use Dominated Convergence Theorem (Theorem 10.4.4) instead of Fatou’s Lemma (Theorem 10.4.3) to get

$$\int_a^b f' d\mu = \lim_{n \rightarrow \infty} \int_a^b \frac{f(x+\epsilon_n) - f(x)}{\epsilon_n} d\mu = \lim_{n \rightarrow \infty} \frac{1}{\epsilon_n} \left( \int_b^{b+\epsilon_n} - \int_a^{a+\epsilon_n} \right) f d\mu.$$

The right side is equal to  $f(b) - f(a)$  by the continuity of  $f$ .

For possibly unbounded  $g \geq 0$ , consider the truncation

$$g_n = \begin{cases} g, & \text{if } 0 \leq g \leq n, \\ n, & \text{if } g > n, \end{cases} \quad f_n(x) = f(a) + \int_a^x g_n d\mu, \quad h_n(x) = \int_a^x (g - g_n) d\mu.$$

Both  $f_n$  and  $h_n$  are increasing functions and are therefore differentiable almost everywhere, with  $f'_n \geq 0$  and  $h'_n \geq 0$ . Since  $g_n$  is bounded, we already proved the integral of  $f'_n$  is  $f_n$ . Then by Theorem 12.3.3 and  $f' = f'_n + h'_n \geq f'_n$ , we have

$$f(b) - f(a) \geq \int_a^b f' d\mu \geq \int_a^b f'_n d\mu = f_n(b) - f_n(a).$$

By the definition of the Lebesgue integral of unbounded measurable functions, we have

$$\lim_{n \rightarrow \infty} (f_n(b) - f_n(a)) = \lim_{n \rightarrow \infty} \int_a^b g_n d\mu = \int_a^b g d\mu = f(b) - f(a).$$

Therefore we conclude that  $\int_a^b f' d\mu = f(b) - f(a)$ .

For general unbound but integrable  $g$ , we may carry out the argument above for  $g^+ = \max\{g, 0\}$  and  $g^- = -\min\{g, 0\}$ . Then combining the two equalities gives us  $\int_a^b f' d\mu = f(b) - f(a)$ .

In the argument above,  $b$  can be replaced by any  $x$  in the interval (if  $x < a$ , some signs and directions of inequalities may need to be changed). So we get  $f(x) = f(a) + \int_a^x f' d\mu$  and  $f(x) = f(a) + \int_a^x g d\mu$  for any  $x$ . By Exercise 10.30, this implies  $f' = g$  almost everywhere.  $\square$

**Exercise 12.30 (Lebesgue Decomposition of Lebesgue-Stieltjes Measure).** Let  $f$  be an increasing function. Let  $f_1(x) = \int_a^x f' d\mu$  for the usual Lebesgue measure  $\mu$  and  $f_0 = f - f_1$ .

1. Prove that  $f_0$  and  $f_1$  are increasing functions. They induce Lebesgue-Stieltjes measures  $\mu_{f_0}$  and  $\mu_{f_1}$ .
2. Let  $X_1 = \{x: f'_0(x) = 0\}$  and  $X_0 = \mathbb{R} - X_1$ . Prove that  $\mathbb{R} = X_0 \sqcup X_1$  is a decomposition that gives  $\mu_{f_0} \perp \mu$ .
3. Prove that  $\mu_f = \mu_{f_0} + \mu_{f_1}$  is the Lebesgue decomposition of  $\mu_f$ .

**Exercise 12.31.** Extend Exercise 12.30 to bounded variation functions.

**Exercise 12.32.** Use Exercises 12.30 and 12.31 to show that  $f(x) = f(a) + \int_a^x f' d\mu$  for absolutely continuous  $f$ .

## 12.4 Differentiation on $\mathbb{R}^n$ : Change of Variable

### Change of Variable

Let  $(X, \Sigma_X, \mu_X)$  and  $(Y, \Sigma_Y, \mu_Y)$  be measure spaces. Let  $\Phi: X \rightarrow Y$  be an invertible map preserving the measurability:  $A \in \Sigma_X$  if and only if  $\Phi(A) \in \Sigma_Y$ . Then for any measurable function  $f$  on  $Y$ ,  $f \circ \Phi$  is a measurable function on  $X$ . The relation between the integral of  $f$  on  $Y$  and the integral of  $f \circ \Phi$  on  $X$  is the change of variable formula.

Since  $\Phi$  is invertible, it maps disjoint union to disjoint union. Then it is easy to see that  $\mu_Y(\Phi(A))$  is a measure on  $X$ . In fact,  $\Phi: (X, \Sigma_X, \mu_Y \circ \Phi) \rightarrow (Y, \Sigma_Y, \mu_Y)$  is an “isomorphism” of measure spaces. This implies that

$$\int_{\Phi(A)} f d\mu_Y = \int_A (f \circ \Phi) d(\mu_Y \circ \Phi).$$

If  $\Phi$  maps subsets of  $\mu_X$ -measure zero to subsets of  $\mu_Y$ -measure zero, then  $\mu_Y \circ \Phi$  is absolutely continuous with respect to  $\mu_X$ . If we also know  $\mu_X$  is  $\sigma$ -finite, then by the Radon-Nikodym Theorem,

$$\mu_Y(\Phi(A)) = \int_A J_\Phi d\mu_X, \quad J_\Phi = \frac{d(\mu_Y \circ \Phi)}{d\mu_X}.$$

The function  $J_\Phi$  is the *Jacobian* of  $\Phi$  and is unique up to a subset of  $\mu_X$ -measure zero.

If we further know that the measure  $\nu = \mu_Y \circ \Phi$  is also  $\sigma$ -finite (by Exercise 12.2, this happens when  $J_\Phi$  does not take infinite value), then Proposition 12.1.8 implies the second equality below

$$\int_{\Phi(A)} f d\mu_Y = \int_A (f \circ \Phi) d(\mu_Y \circ \Phi) = \int_A (f \circ \Phi) J_\Phi d\mu_X.$$

This is the general change of variable formula.

**Proposition 12.4.1 (Change of Variable).** *Suppose  $(X, \Sigma_X, \mu_X)$  and  $(Y, \Sigma_Y, \mu_Y)$  are measure spaces and  $\Phi: X \rightarrow Y$  is an invertible map preserving the measurability:  $A \in \Sigma_X$  if and only if  $\Phi(A) \in \Sigma_Y$ . Suppose  $\mu_X$  is  $\sigma$ -finite, and  $\mu_X(A) = 0$  implies  $\mu_Y(\Phi(A)) = 0$ , so that the Jacobian  $J_\Phi$  exists. If the Jacobian does not take infinite value, then*

$$\int_{\Phi(A)} f d\mu_Y = \int_A (f \circ \Phi) J_\Phi d\mu_X.$$

**Example 12.4.1.** Consider a strictly increasing continuous function  $\alpha: [a, b] \rightarrow [\alpha(a), \alpha(b)]$ , where both intervals have the usual Lebesgue measure. Since both  $\alpha$  and  $\alpha^{-1}$  are continuous, the map  $\alpha$  preserves the Borel measurability. For Borel subsets  $A$ , show that the measure  $\mu(\alpha(A))$  is the Lebesgue-Stieltjes measure  $\mu_\alpha(A)$ .

By Theorem 11.2.2, the key is to show the “regular property” for  $\mu(\alpha(A))$ . For a Borel set  $A$ ,  $\alpha(A)$  is the preimage of  $A$  under continuous map  $\alpha^{-1}$  and is therefore also a Borel set. Thus for any  $\epsilon > 0$ , we have open  $U$  and closed  $C$ , such that  $C \subset \alpha(A) \subset U$  and  $\mu(\alpha(A)) < \epsilon$ . Then  $\alpha^{-1}(C)$  is closed and  $\alpha^{-1}(U)$  is open,  $\alpha^{-1}(C) \subset A \subset \alpha^{-1}(U)$ , such that  $\mu(\alpha(\alpha^{-1}(U) - \alpha^{-1}(C))) = \mu(U - C) < \epsilon$ . This verifies the regularity property in the second statement in Theorem 11.2.2.

Now in case  $\alpha$  is absolutely continuous, by the Fundamental Theorem of Lebesgue integral (Theorem 12.3.5), we have  $J_\alpha = \frac{d\mu_\alpha}{d\mu} = \alpha'$  and

$$\int_{\alpha(a)}^{\alpha(b)} f d\mu = \int_a^b f d\mu_\alpha = \int_a^b f \alpha' d\mu.$$

You may compare the formula with Theorem 4.5.2.

**Example 12.4.2.** For an invertible linear transform  $L: \mathbb{R}^n \rightarrow \mathbb{R}^n$ . Proposition 11.4.5 says  $\mu(L(A)) = |\det(L)|\mu(A)$ . Therefore  $J_L = |\det(L)|$  and we have the linear change of variable formula

$$\int_{L(A)} f d\mu = |\det L| \int_A (f \circ L) d\mu.$$

**Exercise 12.33.** Suppose  $\alpha: [a, b] \rightarrow [\alpha(a), \alpha(b)]$  is an increasing continuous function. Let  $[a_i, b_i]$  be the maximal intervals in  $[a, b]$  such that the restrictions  $\alpha|_{[a_i, b_i]} = c_i$  are constants. Let  $X = [a, b] - \cup [a_i, b_i]$  and  $Y = [\alpha(a), \alpha(b)] - \{c_i\}$ .

1. Prove that the restriction  $\hat{\alpha}: X \rightarrow Y$  is an invertible continuous map, such that the inverse is also continuous.
2. Prove that  $\mu(\alpha(A)) = \mu(\hat{\alpha}(A \cap X))$ .
3. Use the idea of Example 12.4.1 to show that  $\mu(\alpha(A))$  is a regular measure for Borel measurable  $A$ .
4. Prove that  $\mu(\alpha(A)) = \mu_\alpha(A)$ .

What if  $\alpha$  is not assumed continuous?

The two examples suggest what we may expect for the change of variable for the Lebesgue integral on  $\mathbb{R}^n$ . For a map  $\Phi: X \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ , by Proposition 12.2.1 and the remark afterwards, we have

$$J_\Phi = \frac{d(\mu \circ \Phi)}{d\mu} = \lim_{\epsilon \rightarrow 0} \frac{\mu(\Phi(B(\vec{a}, \epsilon)))}{\mu(B(\vec{a}, \epsilon))} \text{ for almost all } \vec{a} \in X. \quad (12.4.1)$$

If  $\Phi$  is differentiable at  $\vec{a}$ , then  $\Phi$  is approximated by the linear map  $F(\vec{x}) = \Phi(\vec{a}) + \Phi'(\vec{a})(\vec{x} - \vec{a})$ , and we expect

$$J_\Phi(\vec{a}) = \lim_{\epsilon \rightarrow 0} \frac{\mu(\Phi(B(\vec{a}, \epsilon)))}{\mu(B(\vec{a}, \epsilon))} = \lim_{\epsilon \rightarrow 0} \frac{\mu(F(B(\vec{a}, \epsilon)))}{\mu(B(\vec{a}, \epsilon))} = |\det \Phi'(\vec{a})|.$$

The third equality is due to the translation invariance of the Lebesgue measure and Proposition 11.4.5. The second equality is based on intuition and needs to be rigorously proved. Then we get the change of variable formula on the Euclidean space

$$\int_{\Phi(A)} f d\mu = \int_A (f \circ \Phi) |\det \Phi'| d\mu.$$

The equality will be rigorously established in Theorem 12.4.5.

## Differentiability of Lipschitz Map

While absolutely continuous functions are enough for the change of variable on  $\mathbb{R}$ , similar role is played by Lipschitz maps on  $\mathbb{R}^n$ .

A map  $\Phi: X \rightarrow Y$  is *Lipschitz* if there is a constant  $L$ , such that

$$\|\Phi(\vec{x}) - \Phi(\vec{x}')\| \leq L\|\vec{x} - \vec{x}'\| \text{ for any } \vec{x}, \vec{x}' \in X.$$

By the equivalence of norms, the definition is independent of the choice of norms, although the constant  $L$  may need to be modified for different choices.

A map  $\Phi$  is *bi-Lipschitz* if there are constants  $L, L' > 0$ , such that

$$L'\|\vec{x} - \vec{x}'\| \leq \|\Phi(\vec{x}) - \Phi(\vec{x}')\| \leq L\|\vec{x} - \vec{x}'\| \text{ for any } \vec{x}, \vec{x}' \in X.$$

This means that  $\Phi: X \rightarrow Y = \Phi(X)$  is invertible, and both  $\Phi$  and  $\Phi^{-1}$  are Lipschitz. The single variable bi-Lipschitz functions already appeared in Exercise 4.11.

A map is Lipschitz if and only if its coordinate functions are Lipschitz. For a Lipschitz function  $f$  on  $X \subset \mathbb{R}^n$  and any  $\vec{a} \in X$ ,

$$f_{\vec{a}}(\vec{x}) = f(\vec{a}) + L\|\vec{x} - \vec{a}\|$$

is a Lipschitz function on the whole  $\mathbb{R}^n$ . It is not hard to see that the infimum of these Lipschitz functions

$$\tilde{f}(\vec{x}) = \inf_{\vec{a} \in X} f_{\vec{a}}(\vec{x})$$

is a Lipschitz function on the whole  $\mathbb{R}^n$  that extends  $f$  on  $X$ . Therefore any Lipschitz map on  $X$  can be extended to a Lipschitz map on  $\mathbb{R}^n$ .

**Proposition 12.4.2.** *A Lipschitz map  $\Phi: X \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  maps Lebesgue measurable subsets to Lebesgue measurable subsets.*

The proposition can be proved just like the first part of the proof of Proposition 11.4.5. The result implies that bi-Lipschitz maps indeed satisfy conditions of Proposition 12.4.1.

**Exercise 12.34.** Prove that a map on a ball (or more generally, a convex subset) with bounded partial derivatives is a Lipschitz map. Moreover, explain that it is possible for a differentiable map to have bounded partial derivatives on an open subset, yet the map is not Lipschitz.

**Exercise 12.35.** Use Theorem 11.4.6 to prove that, if  $\Phi: X \rightarrow Y$  is an invertible map between Lebesgue measurable subsets of  $\mathbb{R}^n$ , such that  $A$  is Lebesgue measurable implies  $\Phi(A)$  is Lebesgue measurable, then  $\mu(A) = 0$  implies  $\mu(\Phi(A)) = 0$ .

To get the formula for a bi-Lipschitz change of variable, we also need to know the differentiability. For absolutely continuous functions on the real line, this was given by Theorem 12.3.3. For Lipschitz functions on the Euclidean space, this is given by the following.

**Theorem 12.4.3 (Rademacher).** *A Lipschitz map is differentiable almost everywhere.*

*Proof.* It is sufficient to prove for each coordinate function of the Lipschitz map. Moreover, we may assume that the Lipschitz function is defined on  $\mathbb{R}^n$ . Note that since a single variable Lipschitz function is absolutely continuous, which implies bounded variation, the Rademacher theorem holds on the real line by Theorem 12.3.3.

We fix a vector  $\vec{v}$  of unit length and consider the subset  $X$  of all  $\vec{x} \in \mathbb{R}^n$ , such that the limit for directional derivative diverges

$$D_{\vec{v}}f(\vec{x}) = \lim_{t \rightarrow 0} \frac{1}{t}(f(\vec{x} + t\vec{v}) - f(\vec{x})).$$

Since  $\frac{1}{t}(f(\vec{x} + t\vec{v}) - f(\vec{x}))$  is continuous for each fixed  $t$ , Example 11.4.4 shows that  $X$  is a  $G_{\delta\sigma}$ -set and is therefore Lebesgue measurable. For the special case  $\vec{v} = \vec{e}_1$  and  $\vec{x} = (x, \vec{y}) \in \mathbb{R} \times \mathbb{R}^{n-1}$ ,  $D_{\vec{e}_1}f(\vec{x}) = \frac{\partial f(x, \vec{y})}{\partial x}$  is the partial derivative. If we fix  $\vec{y}$ , then for the single variable Lipschitz function  $f_{\vec{y}}(x) = f(x, \vec{y})$ , the partial derivative  $\frac{\partial f(x, \vec{y})}{\partial x}$  exists almost everywhere. This means that  $\mu_1(X \cap \mathbb{R} \times \vec{y}) = 0$  for the 1-dimensional Lebesgue measure  $\mu_1$ . The measurability of  $X$  then allows us to apply Proposition 11.3.3 to conclude that the  $n$ -dimensional Lebesgue measure

$$\mu_n(X) = \int_{\mathbb{R}^{n-1}} \mu_1(X \cap \mathbb{R} \times \vec{y}) d\mu_{n-1}(\vec{y}) = \int_{\mathbb{R}^{n-1}} 0 d\mu_{n-1}(\vec{y}) = 0.$$

For general  $\vec{v}$ , we may replace the  $x$ -coordinate direction  $\mathbb{R} \times \vec{0}$  by  $\mathbb{R}\vec{v}$  and replace the  $\vec{y}$ -coordinate direction  $\vec{0} \times \mathbb{R}^{n-1}$  by the subspace orthogonal to  $\vec{v}$  and still conclude that the corresponding non-differentiable subset  $X$  has zero measure.

Now each of the  $n$  partial derivatives exists almost everywhere. Therefore the gradient  $\nabla f = (D_{\vec{e}_1}f, \dots, D_{\vec{e}_n}f)$  exists almost everywhere. On the other hand, for any fixed  $\vec{v}$ , the directional derivative  $D_{\vec{v}}f$  exists almost everywhere. We claim that, for fixed  $\vec{v}$ , we have

$$D_{\vec{v}}f = \nabla f \cdot \vec{v} \quad (12.4.2)$$

almost everywhere. By Proposition 12.2.4, we only need to prove that

$$\int (D_{\vec{v}}f)g d\mu_n = \int (\nabla f \cdot \vec{v})g d\mu_n$$

for any compactly supported smooth function  $g$ .

For the special case  $\vec{v} = \vec{e}_1$ , the Fubini Theorem (Theorem 11.3.4) tells us

$$\int (D_{\vec{e}_1}f)g d\mu_n = \int_{\mathbb{R}^{n-1}} \left( \int_{\mathbb{R}} \frac{\partial f(x, \vec{y})}{\partial x} g(x, \vec{y}) d\mu_1(x) \right) d\mu_{n-1}(\vec{y}).$$

The Lipschitz function  $f_{\vec{y}}(x) = f(x, \vec{y})$  is absolutely continuous, and  $g_{\vec{y}}(x) = g(x, \vec{y})$  is differentiable and bounded. Therefore  $f_{\vec{y}}(x)g_{\vec{y}}(x)$  is absolutely continuous and

$$(f_{\vec{y}}(x)g_{\vec{y}}(x))' = f'_{\vec{y}}(x)g_{\vec{y}}(x) + f_{\vec{y}}(x)g'_{\vec{y}}(x)$$

for almost all  $x$ , and we may apply the fundamental theorem (Theorem 12.3.5) to get integration by parts formula

$$f_{\vec{y}}(b)g_{\vec{y}}(b) - f_{\vec{y}}(a)g_{\vec{y}}(a) = \int_a^b f'_{\vec{y}}(x)g_{\vec{y}}(x) d\mu_1(x) + \int_a^b f_{\vec{y}}(x)g'_{\vec{y}}(x) d\mu_1(x).$$

Since  $g_{\vec{y}}(x) = 0$  for sufficiently big  $x$ , by taking  $a \rightarrow -\infty$  and  $b \rightarrow +\infty$ , the left side is 0, and we get

$$\begin{aligned} \int (D_{\vec{e}_1}f)g d\mu_n &= \int_{\mathbb{R}^{n-1}} \left( \int_{\mathbb{R}} f'_{\vec{y}}(x)g_{\vec{y}}(x) d\mu_1(x) \right) d\mu_{n-1}(\vec{y}) \\ &= - \int_{\mathbb{R}^{n-1}} \left( \int_{\mathbb{R}} f_{\vec{y}}(x)g'_{\vec{y}}(x) d\mu_1(x) \right) d\mu_{n-1}(\vec{y}) = - \int f(D_{\vec{e}_1}g) d\mu_n. \end{aligned}$$

Similarly, for general  $\vec{v}$ , we have

$$\int (D_{\vec{v}}f)g d\mu_n = - \int f(D_{\vec{v}}g) d\mu_n. \quad (12.4.3)$$

Then for  $\vec{v} = (v_1, \dots, v_n)$ , we have

$$\begin{aligned} \int (\nabla f \cdot \vec{v})g d\mu_n &= \int \left( \sum_{i=1}^n v_i D_{\vec{e}_i} f \right) g d\mu_n = \sum_{i=1}^n v_i \int (D_{\vec{e}_i} f)g d\mu_n \\ &= - \sum_{i=1}^n v_i \int f(D_{\vec{e}_i} g) d\mu_n = - \int f \left( \sum_{i=1}^n v_i D_{\vec{e}_i} g \right) d\mu_n \\ &= - \int f(\nabla g \cdot \vec{v}) d\mu_n. \end{aligned} \quad (12.4.4)$$

Since  $g$  is differentiable, the right sides of (12.4.3) and (12.4.4) are equal. Therefore

$$\int (D_{\vec{v}}f)g d\mu_n = \int (\nabla f \cdot \vec{v})g d\mu_n \text{ for any compactly supported smooth } g.$$

Now let  $V$  be a countable dense subset of the unit sphere. For each  $\vec{v} \in V$ , the subset  $Y(\vec{v})$  on which the equality (12.4.2) fails has measure zero. Since  $V$  is countable, the union  $\cup_{\vec{v} \in V} Y(\vec{v})$  still has measure zero. Then at any  $\vec{x} \in Z = \mathbb{R}^n - \cup_{\vec{v} \in V} Y(\vec{v})$ , the equality (12.4.2) holds for all  $\vec{v} \in V$ .

Now we claim that  $f$  is differentiable at any point  $\vec{x} \in Z$ . So we consider

$$\epsilon(t, \vec{v}, \vec{x}) = \frac{1}{t} (f(\vec{x} + t\vec{v}) - f(\vec{x})) - \nabla f \cdot \vec{v}, \quad \vec{x} \in Z, \|\vec{v}\| = 1.$$

The problem is to show that, for any  $\epsilon > 0$ , there is  $\delta > 0$  depending on  $\epsilon, \vec{x}$  but not on  $\vec{v}$ , such that  $|t| < \delta$  implies  $|\epsilon(t, \vec{v}, \vec{x})| < \epsilon$ . We note that, due to the Lipschitz property, we have

$$\left| \frac{1}{t} (f(\vec{x} + t\vec{u}) - f(\vec{x} + t\vec{v})) \right| \leq \frac{1}{|t|} L \|t\vec{u} - t\vec{v}\| = L \|\vec{u} - \vec{v}\|,$$

and if the norm is the Euclidean norm,

$$\|\nabla f \cdot \vec{u} - \nabla f \cdot \vec{v}\| \leq \|\nabla f\| \|\vec{u} - \vec{v}\| \leq \sqrt{n}L \|\vec{u} - \vec{v}\|.$$

Therefore

$$\begin{aligned} |\epsilon(t, \vec{u}, \vec{x}) - \epsilon(t, \vec{v}, \vec{x})| &= \left| \frac{1}{t} (f(\vec{x} + t\vec{u}) - f(\vec{x} + t\vec{v})) - (\nabla f \cdot \vec{u} - \nabla f \cdot \vec{v}) \right| \\ &\leq (1 + \sqrt{n})L \|\vec{u} - \vec{v}\|. \end{aligned}$$

Since  $V$  is dense in the unit sphere, there are finitely many  $\vec{v}_1, \dots, \vec{v}_k \in V$ , such that for any  $\vec{v}$  of unit length, we have  $\|\vec{v} - \vec{v}_i\| < \epsilon$  for some  $i$ . For these  $k$  directions, we can find  $\delta$ , such that  $|t| < \delta$  implies  $|\epsilon(t, \vec{v}_i, \vec{x})| < \epsilon$  for all  $i = 1, \dots, k$ . Then for any  $\vec{v}$  of unit length, we have  $\|\vec{v} - \vec{v}_i\| < \epsilon$  for some  $i$ , and

$$\begin{aligned} |\epsilon(t, \vec{v}, \vec{x})| &\leq |\epsilon(t, \vec{v}_i, \vec{x}) - \epsilon(t, \vec{v}, \vec{x})| + |\epsilon(t, \vec{v}_i, \vec{x})| \\ &\leq (1 + \sqrt{n})L \|\vec{v}_i - \vec{v}\| + |\epsilon(t, \vec{v}_i, \vec{x})| \\ &\leq (1 + \sqrt{n})L\epsilon + \epsilon = [(1 + \sqrt{n})L + 1]\epsilon. \end{aligned}$$



This completes the proof.  $\square$

A remarkable result related to Rademacher's theorem is the following.

**Theorem 12.4.4** (Alexandrov). *A convex function is twice differentiable almost everywhere.*

### Change of Variable on $\mathbb{R}^n$

**Theorem 12.4.5.** *Suppose  $\Phi: X \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a bi-Lipschitz map. Then for a Lebesgue integrable function  $f$  on  $\Phi(X)$ , we have*

$$\int_{\Phi(X)} f d\mu = \int_X f \circ \Phi |\det \Phi'| d\mu.$$

The derivative  $\Phi'$  in the formula comes from the Rademacher theorem (Theorem 12.4.3) and the fact that  $\Phi$  can be extended to a Lipschitz map on  $\mathbb{R}^n$ .

*Proof.* We still denote by  $\Phi$  and  $\Phi^{-1}$  the Lipschitz extensions of  $\Phi$  and  $\Phi^{-1}$  on  $X$  and  $\Phi(X)$ . The two maps are inverse to each other only on  $X$  and  $\Phi(X)$ . By Rademacher theorem, the two (extended) maps  $\Phi$  and  $\Phi^{-1}$  are differentiable away from subsets  $A$  and  $B$  of measure zero. We may also assume that  $A$  contains all the places where the equality (12.4.1) fails, and all places where

$$\lim_{\epsilon \rightarrow 0} \frac{\mu(X \cap B(\vec{a}, \epsilon))}{\mu(B(\vec{a}, \epsilon))} = 1 \quad (12.4.5)$$

fails (see the discussion of Lebesgue density theorem after Proposition 12.2.1). We may further assume that  $B$  contains the similar failure points for  $\Phi^{-1}$ . By Propositions 12.4.1 and 12.4.2,  $\Phi(A)$  and  $\Phi^{-1}(B)$  are subsets of measure zero. Then  $\Phi$  and  $\Phi^{-1}$  are differentiable away from  $A \cup \Phi^{-1}(B)$  and  $\Phi(A) \cup B$ , both being subsets of measure zero. By replacing  $X$  with  $X - A \cup \Phi^{-1}(B)$  (which does not affect the change of variable formula we want to prove, and does not affect the equalities (12.4.1) and (12.4.5)), we may assume that  $\Phi$  is differentiable everywhere in  $X$ , and the equalities (12.4.1) and (12.4.5) hold for every  $\vec{a} \in X$ . Moreover, we may also assume the same happens to  $\Phi^{-1}$  on  $\Phi(X)$ .

We expect the chain rule to tell us  $(\Phi^{-1})'(\Phi(\vec{a})) \circ \Phi'(\vec{a}) = id$  for  $\vec{a} \in X$ . While the claim is true, the reason is a subtle one. The problem is that we only have  $\Phi^{-1} \circ \Phi = id$  on  $X$ , so that  $\Phi^{-1} \circ \Phi = id$  may not hold on a whole neighborhood of  $\vec{a}$ . So we need to repeat the argument for the chain rule again, only over  $X$ . Let  $\vec{b} = \Phi(\vec{a})$ . Then  $\vec{a} = \Phi^{-1}(\vec{b})$ . By the differentiability of (extended)  $\Phi$  and  $\Phi^{-1}$  at  $\vec{a}$  and  $\vec{b}$ , we have

$$\Phi(\vec{x}) = \vec{b} + \Phi'(\vec{a})\Delta\vec{x} + o(\Delta\vec{x}), \quad \Phi^{-1}(\vec{y}) = \vec{a} + (\Phi^{-1})'(\vec{b})\Delta\vec{y} + o(\Delta\vec{y}).$$

Then for  $\vec{x} \in X$ , we have

$$\begin{aligned} \vec{x} &= \Phi^{-1}(\Phi(\vec{x})) = \vec{a} + (\Phi^{-1})'(\vec{b})(\Phi'(\vec{a})\Delta\vec{x} + o(\Delta\vec{x})) + o(\Phi'(\vec{a})\Delta\vec{x} + o(\Delta\vec{x})) \\ &= \vec{a} + (\Phi^{-1})'(\vec{b})\Phi'(\vec{a})\Delta\vec{x} + o(\Delta\vec{x}). \end{aligned}$$

In other words,  $\vec{v} = (\Phi^{-1})'(\vec{b})\Phi'(\vec{a})\vec{v} + o(\vec{v})$  as long as  $\vec{a} + \vec{v} \in X$ . Since (12.4.5) is assumed to hold at  $\vec{a}$ , there are plenty enough small  $\vec{v}$  pointing to all directions, such that  $\vec{a} + \vec{v} \in X$ . Then the equality  $\vec{v} = (\Phi^{-1})'(\vec{b})\Phi'(\vec{a})\vec{v} + o(\vec{v})$  for these  $\vec{v}$  implies that  $(\Phi^{-1})'(\vec{b})\Phi'(\vec{a}) = id$ . The rigorous proof is given by Proposition 12.4.6.

Now  $(\Phi^{-1})'(\Phi(\vec{a}))\Phi'(\vec{a}) = id$  holds for every  $\vec{a} \in X$ . This implies that  $\Phi'$  is invertible on  $X$  and  $(\Phi^{-1})'$  is invertible on  $\Phi(X)$ .

Let  $\vec{a} \in X$ ,  $\vec{b} = \Phi(\vec{a})$  and  $K = \Phi'(\vec{a})^{-1}$ . Then for any  $\epsilon > 0$ , there is  $\delta' > 0$ , such that

$$\begin{aligned} \|\vec{x} - \vec{a}\| < \delta' &\implies \|\Phi(\vec{x}) - \Phi(\vec{a}) - \Phi'(\vec{a})(\vec{x} - \vec{a})\| \leq \epsilon\|\vec{x} - \vec{a}\| \\ &\implies \|K(\Phi(\vec{x})) - K(\vec{b}) - (\vec{x} - \vec{a})\| \leq \|K\|\epsilon\|\vec{x} - \vec{a}\| \\ &\implies \|K(\Phi(\vec{x})) - K(\vec{b})\| \leq (1 + \|K\|\epsilon)\|\vec{x} - \vec{a}\|. \end{aligned}$$

This implies

$$0 < \delta \leq \delta' \implies K(\Phi(B(\vec{a}, \delta))) \subset B(K(\vec{b}), (1 + \|K\|\epsilon)\delta).$$

Therefore for  $0 < \delta \leq \delta'$ , we have

$$\begin{aligned} \frac{\mu(\Phi(B(\vec{a}, \delta)))}{\mu(B(\vec{a}, \delta))} &= \frac{1}{|\det K|} \frac{\mu(K(\Phi(B(\vec{a}, \delta))))}{\mu(B(K(\vec{b}), \delta))} \\ &\leq |\det \Phi'(\vec{a})| \frac{\mu(B(K(\vec{b}), (1 + \|K\|\epsilon)\delta))}{\mu(B(K(\vec{b}), \delta))} \\ &= |\det \Phi'(\vec{a})|(1 + \|K\|\epsilon)^n. \end{aligned}$$

By our assumption, the limit (12.4.1) converges, and we get

$$J_\Phi(\vec{a}) = \lim_{\delta \rightarrow 0} \frac{\mu(\Phi(B(\vec{a}, \delta)))}{\mu(B(\vec{a}, \delta))} \leq |\det \Phi'(\vec{a})|(1 + \|K\|\epsilon)^n.$$

Since  $\epsilon$  is arbitrary, we get

$$J_\Phi(\vec{a}) \leq |\det \Phi'(\vec{a})| \text{ for } \vec{a} \in X.$$

By applying the same argument to  $\Phi^{-1}$  on  $\Phi(X)$ , we get

$$J_{\Phi^{-1}}(\vec{b}) \leq |\det(\Phi^{-1})'(\vec{b})| = \frac{1}{|\det \Phi'(\vec{a})|} \text{ for } \vec{b} = \Phi(\vec{a}), \vec{a} \in X.$$

However, applying the change  $\Phi$  and then the change  $\Phi^{-1}$  should get back the original integral, so that

$$J_{\Phi^{-1}}(\vec{b})J_\Phi(\vec{a}) = 1 \text{ for almost all } \vec{a} \in X.$$

Therefore we conclude that  $J_\Phi = |\det \Phi'|$  almost everywhere.  $\square$

By taking the linear transform to be  $(\Phi^{-1})'(\vec{b})\Phi'(\vec{a}) - I$ , the following gives rigorous argument for  $(\Phi^{-1})'(\vec{b})\Phi'(\vec{a}) = I$  in the proof above.

**Proposition 12.4.6.** Suppose  $X \subset \mathbb{R}^n$  is a Lebesgue measurable subset satisfying

$$\lim_{\epsilon \rightarrow 0} \frac{\mu(X \cap B_\epsilon)}{\mu(B_\epsilon)} = 1, \quad B_\epsilon = B(\vec{0}, \epsilon).$$

Suppose  $L$  is a linear transform satisfying

$$\lim_{\vec{x} \in X, \vec{x} \rightarrow \vec{0}} \frac{\|L(\vec{x})\|}{\|\vec{x}\|} = 0.$$

Then  $L$  is the zero transform.

*Proof.* Suppose  $L$  is not the zero transform. Then  $\|L(\vec{v})\| > 0$  for some  $\vec{v}$ . By the continuity of  $\lambda(\vec{x}) = \frac{\|L(\vec{x})\|}{\|\vec{x}\|}$  for nonzero  $\vec{x}$ , we have  $\lambda(\vec{x}) > c = \frac{1}{2}\lambda(\vec{v})$  for all  $\vec{x}$  in a ball  $B = B(\vec{v}, \delta)$ . Moreover, for any vector in the open cone  $C = \{t\vec{x} : t \neq 0, \vec{x} \in B\}$ , we also have  $\lambda(t\vec{x}) = \lambda(\vec{x}) > c$ . On the other hand, the assumption  $\lim_{\vec{x} \in X, \vec{x} \rightarrow \vec{0}} \lambda(\vec{x}) = 0$  tells us that there is  $\delta > 0$ , such that  $\lambda(\vec{x}) < c$  for  $\vec{x} \in X \cap B_\delta$ . Therefore  $C \cap X \cap B_\delta = \emptyset$ . This implies that, for  $\epsilon < \delta$ , we have  $X \cap B_\epsilon \subset B_\epsilon - C$ , so that

$$\frac{\mu(X \cap B_\epsilon)}{\mu(B_\epsilon)} \leq \frac{\mu(B_\epsilon - C)}{\mu(B_\epsilon)}.$$

The ratio on the right is independent of  $\epsilon$  and is  $< 1$ . Therefore we cannot have  $\lim_{\epsilon \rightarrow 0} \frac{\mu(X \cap B_\epsilon)}{\mu(B_\epsilon)} = 1$ .  $\square$

## 12.5 Additional Exercise

### Alternative Proof of Radon-Nikodym Theorem

The following is a proof the Radon-Nikodym Theorem (Theorem 12.1.6) for the case  $\mu$  and  $\nu$  are finite non-negative valued measures. The idea is that the function  $f$  should be the “upper envelope” of measurable functions  $f \geq 0$  satisfying

$$\nu(A) \geq \int_A f d\mu \text{ for any measurable } A. \quad (12.5.1)$$

**Exercise 12.36.** Prove that if  $f, g$  satisfy (12.5.1), then  $\max\{f, g\}$  satisfies (12.5.1).

**Exercise 12.37.** Prove that there is an increasing sequence  $f_n$  satisfying (12.5.1), such that

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \sup \left\{ \int_X g d\mu : g \text{ satisfies (12.5.1)} \right\}.$$

Moreover, prove that  $f = \lim_{n \rightarrow \infty} f_n$  also satisfies (12.5.1) and  $\lambda(A) = \nu(A) - \int_A f d\mu$  is a measure.

**Exercise 12.38.** Prove that, if  $\lambda(X) > 0$ , then there is  $\epsilon > 0$  and a measurable  $B$ , such that  $\mu(B) > 0$  and  $f + \epsilon \chi_B$  satisfies (12.5.1).

Exercise 12.39. Use Exercise 12.38 to prove that  $\nu(A) = \int_A f d\mu$  for any measurable  $A$ .

**Chapter 13**

# **Multivariable Integration**

## 13.1 Curve

Curves, surfaces or more generally submanifolds of  $\mathbb{R}^n$  have length, area or volume. These are measures at various dimensions, and the integration may be defined against these measures.

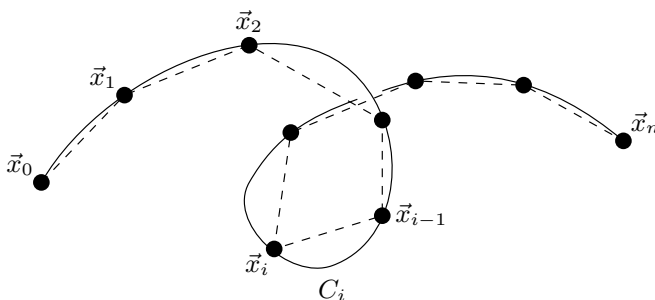
By a *curve*  $C$ , we mean that the curve can be presented by a continuous map  $\phi(t): [a, b] \rightarrow \mathbb{R}^n$ , called a *parameterization* of  $C$ . A change of variable (or *reparameterization*) is an invertible continuous map  $u \in [c, d] \mapsto t = t(u) \in [a, b]$ , so that  $C$  is also presented by  $\phi(t(u)): [c, d] \rightarrow \mathbb{R}^n$ . We take all reparameterizations to be equally good. In other words, we do not have a preferred choice of parameterization.

### Length of Curve

To define the length of a curve  $C$ , we take a *partition*  $P$ . This is a “monotone” choice of partition points  $\vec{x}_i \in C$ . The segment  $C_i$  of  $C$  between  $\vec{x}_{i-1}$  and  $\vec{x}_i$  is approximated by the straight line connecting the two points. The length of the curve is then approximately

$$\mu_P(C) = \sum \|\vec{x}_i - \vec{x}_{i-1}\|.$$

Here we do not have to restrict the norm to the euclidean one.



**Figure 13.1.1.** Partition and approximation of curve.

If another partition  $P'$  refines  $P$ , then by the triangle inequality, we have  $\mu_{P'}(C) \geq \mu_P(C)$ . If  $\mu_P(C)$  is bounded, then the curve is *rectifiable* and has *length*

$$\mu(C) = \sup_P \mu_P(C).$$

Given a parameterization  $\phi: [a, b] \rightarrow \mathbb{R}^n$  of  $C$ , a partition  $Q = \{t_i\}$  of  $[a, b]$  gives a partition  $P = \phi(Q) = \{\vec{x}_i = \phi(t_i)\}$  of  $C$ . Moreover, the segment  $C_i$  may be parameterized by the restriction  $\phi|_{[t_{i-1}, t_i]}$ . Since the definition of length does not make explicit use of the parameterization, the length is independent of the choice of parameterization.

**Proposition 13.1.1.** *A curve is rectifiable if and only if its coordinate functions have bounded variations.*

*Proof.* Because all norms are equivalent, the rectifiability is independent of the choice of the norm. If we take the  $L^1$ -norm, then for a parameterization  $\phi(t)$  and a partition  $Q$  of  $[a, b]$ , we have

$$\|\vec{x}_i - \vec{x}_{i-1}\|_1 = \|\phi(t_i) - \phi(t_{i-1})\|_1 = |x_1(t_i) - x_1(t_{i-1})| + \cdots + |x_n(t_i) - x_n(t_{i-1})|.$$

This implies

$$\mu_P(C) = V_P(x_1) + \cdots + V_P(x_n),$$

and the proposition follows.  $\square$

For a parameterized rectifiable curve  $C$ , the *arc length* function

$$s(t) = \mu(\phi|_{[a,t]})$$

is increasing. By Proposition 4.6.3,  $s(t)$  is also continuous. If  $\phi$  is not constant on any interval in  $[a, b]$ , then  $s(t)$  is strictly increasing, and the curve can be reparameterized by the arc length. In general, we may take out the intervals on which  $\phi$  is constant and modify the parameterization by reducing such intervals to single points. Since this does not effectively change  $C$ , without loss of generality, we may assume that all curves can be reparameterized by the arc length.

**Exercise 13.1.** Prove that the  $L^1$ -length of a rectifiable curve is the sum of the variations of the coordinates on the interval. Then find the  $L^1$ -length of the unit circle in  $\mathbb{R}^2$ .

**Exercise 13.2.** Show that a straight line segment is rectifiable and compute its length.

**Exercise 13.3.** Suppose  $C$  is a rectifiable curve.

1. Prove that if  $C$  is divided into two parts  $C_1$  and  $C_2$ , then  $\mu(C) = \mu(C_1) + \mu(C_2)$ .
2. Prove that if  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$  satisfies  $\|F(\vec{x}) - F(\vec{y})\| = \|\vec{x} - \vec{y}\|$ , then  $\mu(C) = \mu(F(C))$ .
3. Prove that if  $aC$  is obtained by scaling the curve by factor  $a$ , then  $\mu(aC) = |a|\mu(C)$ .

**Exercise 13.4.** Let  $n \geq 2$ . Prove that the image of any rectifiable curve in  $\mathbb{R}^n$  has  $n$ -dimensional measure zero. In fact, the image is Jordan measurable.

**Exercise 13.5.** Let  $C$  be a rectifiable curve. Prove that for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that if a partition  $P = \{C_i\}$  of  $C$  satisfies  $\|P\| = \max_i \mu(C_i) < \delta$ , then  $\mu_P(C) > \mu(C) - \epsilon$ . This implies

$$\lim_{\|P\| \rightarrow 0} \mu_P(C) = \mu(C).$$

**Exercise 13.6.** Suppose  $\phi(t)$  is a parameterization of a rectifiable curve  $C$ . Let  $Q = \{t_i\}$  be a partition of the parameter and  $P = \{\phi(t_i)\}$  be the corresponding partition of the curve.

1. Prove that  $\|Q\| = \max_i |t_i - t_{i-1}| \rightarrow 0$  implies  $\|P\| = \max_i \mu(C_i) \rightarrow 0$ .
2. Suppose  $\phi$  is not constant on any interval. Prove that  $\|P\| \rightarrow 0$  implies  $\|Q\| \rightarrow 0$ .

**Exercise 13.7.** Consider a partition of a rectifiable curve  $C$  consisting of partition points  $\vec{x}_i$  and partition segments  $C_i$ .

1. Prove that  $\max_i \mu(C_i) \rightarrow 0$  implies  $\max_i \|\vec{x}_i - \vec{x}_{i-1}\| \rightarrow 0$ .
2. Suppose  $C$  *simple* in the sense that it does not cross itself. Prove that  $\max_i \|\vec{x}_i - \vec{x}_{i-1}\| \rightarrow 0$  implies  $\max_i \mu(C_i) \rightarrow 0$ .

A parameterization of a curve is *absolutely continuous* if its coordinate functions are absolutely continuous. Exercise 13.12 gives other equivalent and more intrinsic definitions. Such a parameterization has the tangent vector  $\phi'(t)$  for almost all  $t$ , and the tangent vector can be used to compute the length of the curve.

**Proposition 13.1.2.** *An absolutely continuous curve  $\phi(t)$  has arc length  $\int_a^b \|\phi'(t)\| dt$ .*

The proposition implies that the arc length function is

$$s(t) = \int_a^t \|\phi'(\tau)\| d\tau,$$

or

$$s'(t) = \|\phi'(t)\| \quad \text{almost everywhere.}$$

In the special case  $t = s$  is the arc length, we find that the tangent vector  $\phi'(s)$  has the unit length.

*Proof.* The absolute continuity implies  $\phi(t) = \phi(a) + \int_a^t \phi'(\tau) d\tau$ . We need to show

$$\mu_P(\phi) = \sum \|\phi(t_i) - \phi(t_{i-1})\| = \sum \left\| \int_{t_{i-1}}^{t_i} \phi'(t) dt \right\|, \quad P = \phi(Q)$$

converges to  $\int_a^b \|\phi'(t)\| dt$  as the partition  $Q$  of  $[a, b]$  gets more and more refined.

The idea is to reduce the problem to the case  $\phi'$  is Riemann integrable.

By Proposition 10.5.3, for any  $\epsilon > 0$ , there is a compactly supported smooth map  $g: [a, b] \rightarrow \mathbb{R}^n$  (actually Riemann integrable is enough for our purpose), such that  $\int_a^b \|\phi' - g\| dt < \epsilon$ . Let  $\gamma(t) = \phi(a) + \int_a^t g(\tau) d\tau$ . Then by the inequality in Exercise 10.61 (also see Exercise 10.71), we have

$$\begin{aligned} |\mu_P(\phi) - \mu_P(\gamma)| &\leq \sum \left\| \int_{t_{i-1}}^{t_i} (\phi' - g) dt \right\| \\ &\leq \sum \int_{t_{i-1}}^{t_i} \|\phi' - g\| dt = \int_a^b \|\phi' - g\| dt < \epsilon, \\ \left| \int_a^b \|\phi'\| dt - \int_a^b \|g\| dt \right| &= \int_a^b \left| \|\phi'\| - \|g\| \right| dt \leq \int_a^b \|\phi' - g\| dt < \epsilon. \end{aligned}$$



On the other hand, we have

$$\begin{aligned} |\mu_P(\gamma) - S(Q, \|g\|)| &\leq \sum \left\| \int_{t_{i-1}}^{t_i} g(t) dt - \|g(t_i^*)\| \Delta t_i \right\| \\ &\leq \sum \left\| \int_{t_{i-1}}^{t_i} g(t) dt - \Delta t_i g(t_i^*) \right\| \\ &\leq \sum \sup_{t \in [t_{i-1}, t_i]} \|g(t) - g(t_i^*)\| \Delta t_i. \end{aligned}$$

The Riemann integrability of  $g$  implies that the right side is  $< \epsilon$  for sufficiently refined  $Q$ . Then we get

$$\left| \mu_P(\phi) - \int_a^b \|g\| dt \right| \leq |\mu_P(\phi) - S(Q, \|g\|)| + \left| S(Q, \|g\|) - \int_a^b \|g\| dt \right| < 2\epsilon$$

for sufficiently refined  $Q$ . Combining all the inequalities, we get

$$\left| \mu_P(\phi) - \int_a^b \|\phi'\| dt \right| < 4\epsilon$$

for sufficiently refined  $Q$ . This implies that  $\int_a^b \|\phi'(t)\| dt$  is the length of  $\phi(t)$ .  $\square$

**Example 13.1.1.** The graph of a function  $f(x)$  on  $[a, b]$  is parameterized by  $\phi(x) = (x, f(x))$ . If  $f$  is absolutely continuous, then the Euclidean length of the graph is  $\int_a^b \sqrt{1 + f'^2} dt$  and the  $L^1$ -length is  $\int_a^b (1 + |f'|) dt$ .

**Example 13.1.2.** For the circle  $\phi(\theta) = (a \cos \theta, a \sin \theta)$ , the Euclidean arc length (counted from  $\theta = 0$ ) is

$$s = \int_0^\theta \|\phi'(t)\|_2 dt = \int_0^\theta \sqrt{a^2 \sin^2 t + a^2 \cos^2 t} dt = a\theta.$$

Therefore the circle is parameterized as  $\phi(s) = \left(a \cos \frac{s}{a}, a \sin \frac{s}{a}\right)$  by the arc length.

On the other hand, with respect to the  $L^1$ -norm, we have

$$s = \int_0^\theta \|\phi'(t)\|_1 dt = \int_0^\theta |a|(|\sin t| + |\cos t|) dt = |a|(1 - \cos \theta + \sin \theta), \quad \text{for } 0 \leq \theta \leq \frac{\pi}{2}.$$

**Example 13.1.3.** The astroid  $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$  can be parameterized as  $x = a \cos^3 t$ ,  $y = a \sin^3 t$  for  $0 \leq t \leq 2\pi$ . The Euclidean length of the curve is

$$\int_0^{2\pi} \sqrt{(-3a \cos^2 t \sin t)^2 + (3a \sin^2 t \cos t)^2} dt = 6a.$$

The  $L^\infty$ -length is

$$\int_0^{2\pi} \max\{|-3a \cos^2 t \sin t|, |3a \sin^2 t \cos t|\} dt = 8 \left(1 - \frac{1}{2\sqrt{2}}\right) a.$$

**Exercise 13.8.** Find the formula for the arc length of a curve in  $\mathbb{R}^2$  in terms of the parameterized polar coordinate  $r = r(t)$ ,  $\theta = \theta(t)$ .

**Exercise 13.9.** Compute the Euclidean lengths of the curves. Can you also find the length with respect to other norms?

1. Parabola  $y = x^2$ ,  $0 \leq x \leq 1$ .
2. Spiral  $r = a\theta$ ,  $0 \leq \theta \leq \pi$ .
3. Another spiral  $r = e^{a\theta}$ ,  $0 \leq \theta \leq \alpha$ .
4. Cycloid  $x = a(t - \sin t)$ ,  $y = a(1 - \cos t)$ ,  $0 \leq t \leq 2\pi$ .
5. Cardioid  $r = 2a(1 + \cos \theta)$ .
6. Involute of the unit circle  $x = \cos \theta + \theta \sin \theta$ ,  $y = \sin \theta - \theta \cos \theta$ ,  $0 \leq \theta \leq \alpha$ .
7. Helix  $x = a \cos \theta$ ,  $y = a \sin \theta$ ,  $z = b\theta$ ,  $0 \leq \theta \leq \alpha$ .

**Exercise 13.10.** Directly prove Proposition 13.1.2 by using Exercise 11.15 and Theorem 12.3.5.

**Exercise 13.11.** What is the length of the graph  $(t, \kappa(t))$ ,  $t \in [0, 1]$ , of the Cantor function  $\kappa$  in Example 11.4.5? Please consider various norms. Explain why you cannot calculate the length by using the formula in Proposition 13.1.2. Instead, by Exercise 13.6, you can use special partitions to calculate the length.

**Exercise 13.12.** Prove the following are equivalent for a continuous curve  $\phi(t)$ .

1. The coordinate functions of  $\phi(t)$  are absolutely continuous.
2. The arc length function  $s(t)$  is absolutely continuous.
3. For any  $\epsilon > 0$ , there is  $\delta > 0$ , such that for disjoint intervals  $(t_i, t'_i)$ , we have

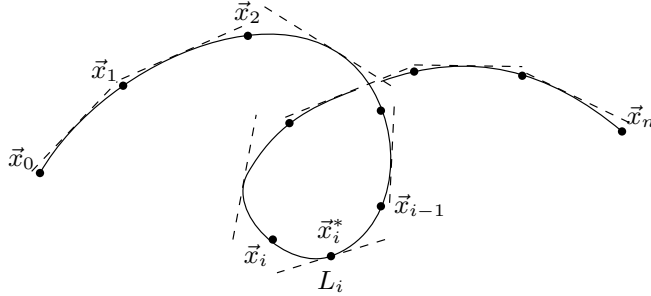
$$|t_1 - t'_1| + \cdots + |t_n - t'_n| < \delta \implies \|\phi(t_1) - \phi(t'_1)\| + \cdots + \|\phi(t_n) - \phi(t'_n)\| < \epsilon.$$

**Exercise 13.13.** Suppose  $\phi(t)$  is a continuously differentiable curve on  $[a, b]$ . For any partition  $Q$  of  $[a, b]$  and choice  $t_i^*$ , the curve is approximated by the tangent lines  $L_i(t) = \phi(t_i^*) + \phi'(t_i^*)(t - t_i^*)$  on intervals  $[t_{i-1}, t_i]$ . See Figure 13.1.2. Prove that the sum of the lengths of the tangent lines converges to the length of  $\phi$  as  $\|Q\| \rightarrow 0$ .

## Integration of Function Along Curve

Let  $f(\vec{x})$  be a function defined along a rectifiable curve  $C$ . For a partition  $P$  of  $C$  and sample points  $\vec{x}_i^* \in C_i$ , we get a Riemann sum

$$S(P, f) = \sum f(\vec{x}_i^*) \mu(C_i).$$



**Figure 13.1.2.** *Tangential approximation of the length of curve.*

If the sum converges as  $\|P\| = \max_i \mu(C_i)$  converges to 0, then the limit is the Riemann integral  $\int_C f ds$  of  $f$  along  $C$ .

If  $C$  is parameterized by  $\phi: [a, b] \rightarrow \mathbb{R}^n$  and  $P$  comes from a partition of  $[a, b]$ , then

$$S(P, f) = \sum f(\phi(t_i^*)) (s(t_i) - s(t_{i-1})).$$

Therefore in case  $f(\phi(t))$  is Riemann-Stieltjes integrable with respect to  $s(t)$ , we have the first equality below

$$\int_C f ds = \int_a^b f(\phi(t)) ds(t) = \int_a^b f(\phi(t)) \|\phi'(t)\| dt.$$

The second equality holds if we further know that  $\phi$  is absolutely continuous.

For maps  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the Riemann integral  $\int_C F ds$  can be similarly defined and computed.

The integral along a curve has properties similar to the Riemann integral on an interval.

**Example 13.1.4.** We try to compute the integral of a linear function  $l(\vec{x}) = \vec{a} \cdot \vec{x}$  along the straight line  $C$  connecting  $\vec{u}$  to  $\vec{v}$ . The straight line can be parameterized as  $\phi(t) = \vec{u} + t(\vec{v} - \vec{u})$ , with  $ds = \|\vec{v} - \vec{u}\| dt$ . Therefore

$$\int_C \vec{a} \cdot \vec{x} ds = \int_0^1 \vec{a} \cdot (\vec{u} + t(\vec{v} - \vec{u})) \|\vec{v} - \vec{u}\| dt = \frac{1}{2} (\vec{a} \cdot (\vec{u} + \vec{v})) \|\vec{v} - \vec{u}\|.$$

The computation holds for any norm.

**Example 13.1.5.** The integral of  $|y|$  along the unit circle with respect to the Euclidean norm is

$$\int_{x^2+y^2=1} |y| ds = \int_0^{2\pi} |\sin \theta| d\theta = 4 \int_0^{\frac{\pi}{2}} |\sin \theta| d\theta = 4.$$

The integral with respect to the  $L^1$ -norms is

$$\int_{x^2+y^2=1} |y| ds = \int_0^{2\pi} |\sin \theta| (|\sin \theta| + |\cos \theta|) d\theta = 4 \int_0^{\frac{\pi}{2}} \sin \theta (\sin \theta + \cos \theta) d\theta = \pi + 2.$$

**Exercise 13.14.** Compute the integral along the curve with respect to the Euclidean norm. Can you also compute the integral with respect to other norms?

1.  $\int_C xy ds$ ,  $C$  is the part of the ellipse  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$  in the first quadrant.
2.  $\int_C (x^{\frac{4}{3}} + y^{\frac{4}{3}}) ds$ ,  $C$  is the astroid  $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$ .
3.  $\int_C (a_1x + a_2y + a_3z) ds$ ,  $C$  is the circle  $x^2 + y^2 + z^2 = 1$ ,  $b_1x + b_2y + b_3z = 0$ .

**Exercise 13.15.** Extend Exercise 13.5 to integral along rectifiable curve

$$\lim_{\max_i \mu(C_i) \rightarrow 0} S(P, f) = \int_C f ds.$$

Then calculate the integral of the function  $y$  along the graph  $(t, \kappa(t))$ ,  $t \in [0, 1]$ , of the Cantor function  $\kappa$  in Exercise 13.11.

**Exercise 13.16.** Prove that  $f$  is Riemann integrable along a rectifiable curve  $C$  if and only if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$  implies  $\sum \omega_{C_i}(f) \mu(C_i) < \epsilon$ . This implies that the Riemann integrability along a rectifiable curve is independent of the choice of the norm.

**Exercise 13.17.** Prove that continuous functions and monotone functions (the concept is defined along curves) are Riemann integrable along rectifiable curves.

**Exercise 13.18.** Suppose  $f$  is Riemann integrable along a rectifiable curve. Suppose  $g$  is a uniformly continuous function on values of  $f$ . Prove that  $g \circ f$  is Riemann integrable along the curve.

**Exercise 13.19.** Prove that the sum and the product of Riemann integrable functions along a rectifiable curve are still Riemann integrable along the curve, and

$$\int_C (f + g) ds = \int_C f ds + \int_C g ds, \quad \int_C c f ds = c \int_C f ds.$$

Moreover, prove that

$$f \leq g \implies \int_C f ds \leq \int_C g ds.$$

**Exercise 13.20.** Suppose  $f$  is a continuous function along a rectifiable curve  $C$ . Prove that there is  $\vec{c} \in C$ , such that  $\int_C f ds = f(\vec{c}) \mu(C)$ .

**Exercise 13.21.** Suppose a rectifiable curve  $C$  is divided into two parts  $C_1$  and  $C_2$ . Prove that a function is Riemann integrable on  $C$  if and only if it is Riemann integrable on  $C_1$  and  $C_2$ . Moreover,

$$\int_C f ds = \int_{C_1} f ds + \int_{C_2} f ds.$$

**Exercise 13.22.** Suppose  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a map satisfying  $\|F(\vec{x}) - F(\vec{y})\| = \|\vec{x} - \vec{y}\|$ . Prove that  $\int_{F(C)} f(\vec{x}) ds = \int_C f(F(\vec{x})) ds$ .

**Exercise 13.23.** Suppose  $f$  is a Riemann integrable function along a rectifiable curve  $C$  connecting  $\vec{a}$  to  $\vec{b}$ . For any  $\vec{x} \in C$ , denote by  $C[\vec{a}, \vec{x}]$  the part of  $C$  between  $\vec{a}$  and  $\vec{x}$ . Then  $F(\vec{x}) = \int_{C[\vec{a}, \vec{x}]} f ds$  can be considered as the “antiderivative” of  $f$  along  $C$ . By using the concept, state and prove the integration by part formula for the integral along  $C$ .

**Exercise 13.24.** Prove that a map  $F: C \rightarrow \mathbb{R}^m$  on a rectifiable curve  $C$  is Riemann integrable if and only if each coordinate of  $F$  is Riemann integrable. Moreover, discuss properties of the integral  $\int_C F ds$ .

### Integration of 1-Form Along Curve

A vector field  $F: C \rightarrow \mathbb{R}^n$  along a rectifiable curve  $C \in \mathbb{R}^n$  assigns a vector  $F(\vec{x}) \in \mathbb{R}^n$  to any point  $\vec{x} \in C$ . For a partition  $P$  of  $C$  and sample points  $\vec{x}_i^* \in C_i$ , define the Riemann sum

$$S(P, F) = \sum F(\vec{x}_i^*) \cdot (\vec{x}_i - \vec{x}_{i-1}).$$

If the sum converges as  $\|P\| = \max_i \mu(C_i)$  converges to 0, then the limit is the Riemann integral  $\int_C F \cdot d\vec{x}$ .

Note that  $\int_C F \cdot d\vec{x}$  depends on the *orientation* of  $C$ , which is the choice of a beginning point  $\vec{x}_0$  (which implies that  $\vec{x}_n$  is the end point). If the direction of the curve is reversed, then the order of the partition points  $\vec{x}_i$  is reversed, and the Riemann sum changes the sign. The observation leads to

$$\int_{-C} F \cdot d\vec{x} = - \int_C F \cdot d\vec{x},$$

where we use  $-C$  to denote the same curve but with reversed orientation. The equality can be compared with  $\int_b^a f dx = - \int_a^b f dx$ .

If  $F = (f_1, \dots, f_n)$ ,  $C$  is parameterized by  $\phi = (x_1, \dots, x_n): [a, b] \rightarrow \mathbb{R}^n$ , and  $P$  comes from a partition of  $[a, b]$ , then

$$S(P, F) = \sum f_1(\phi(t_i^*)) (x_1(t_i) - x_1(t_{i-1})) + \dots + \sum f_n(\phi(t_i^*)) (x_n(t_i) - x_n(t_{i-1})).$$

If  $f_i(\phi(t))$  are Riemann-Stieljes integrable with respect to  $x_i(t)$ , then

$$\int_C F \cdot d\vec{x} = \int_a^b f_1(\phi(t)) dx_1(t) + \dots + \int_a^b f_n(\phi(t)) dx_n(t).$$

If we further know that  $\phi(t)$  is absolutely continuous, then we have

$$\int_C F \cdot d\vec{x} = \int_a^b F(\phi(t)) \cdot \phi'(t) dt.$$

Because of the connection to the Riemann-Stieljes integral, we also denote

$$\int_C F \cdot d\vec{x} = \int_C f_1 dx_1 + \cdots + f_n dx_n.$$

The expression

$$F \cdot d\vec{x} = f_1 dx_1 + \cdots + f_n dx_n$$

is called a 1-form (a *differential form* of order 1).

**Example 13.1.6.** Consider three curves connecting  $(0,0)$  to  $(1,1)$ . The curve  $C_1$  is the straight line  $\phi(t) = (t, t)$ . The curve  $C_2$  is the parabola  $\phi(t) = (t, t^2)$ . The curve  $C_3$  is the straight line from  $(0,0)$  to  $(1,0)$  followed by the straight line from  $(1,0)$  to  $(1,1)$ . Then

$$\begin{aligned} \int_{C_1} ydx + xdy &= \int_0^1 (tdt + tdt) = 1, \\ \int_{C_2} ydx + xdy &= \int_0^1 (t^2 dt + t \cdot 2t dt) = 1, \\ \int_{C_3} ydx + xdy &= \int_0^1 0dx + \int_0^1 1dy = 1. \end{aligned}$$

We note that the result is independent of the choice of the curve. In fact, for any absolutely continuous  $\phi(t) = (x(t), y(t))$ ,  $t \in [a, b]$ , connecting  $(0,0)$  to  $(1,1)$ , we have

$$\begin{aligned} \int_C ydx + xdy &= \int_a^b (y(t)x'(t) + x(t)y'(t))dt = \int_a^b (x(t)y(t))' dt \\ &= x(b)y(b) - x(a)y(a) = 1 - 0 = 1. \end{aligned}$$

**Example 13.1.7.** Taking the three curves in Example 13.1.6 again, we have

$$\begin{aligned} \int_{C_1} xydx + (x+y)dy &= \int_0^1 (t^2 dt + 2t dt) = \frac{4}{3}, \\ \int_{C_2} xydx + (x+y)dy &= \int_0^1 (t^3 dt + (t+t^2)2t dt) = \frac{17}{12}, \\ \int_{C_3} xydx + (x+y)dy &= \int_0^1 0dx + \int_0^1 (1+y)dy = \frac{3}{2}. \end{aligned}$$

In contrast to the integral of  $ydx + xdy$ , the integral of  $xydx + (x+y)dy$  depends on the curve connecting the two points.

**Example 13.1.8.** Let  $F(\vec{x}) = (f(\vec{x}), -f(\vec{x}))$  and let  $C$  be the diagonal connecting  $(0,0)$  to  $(1,1)$ . Then a partition of  $C$  is given by  $\vec{x}_i = (x_i, x_i)$ , and the sample points are  $\vec{x}_i^* = (x_i^*, x_i^*)$ . We have

$$\begin{aligned} S(P, F) &= \sum (f(\vec{x}_i^*), -f(\vec{x}_i^*)) \cdot (\vec{x}_i - \vec{x}_{i-1}) \\ &= \sum (f(x_i^*, x_i^*), -f(x_i^*, x_i^*)) \cdot (x_i - x_{i-1}, x_i - x_{i-1}) \\ &= \sum 0 = 0. \end{aligned}$$

So  $\int_C F \cdot d\vec{x} = 0$  for any function  $f$ . The example shows that, although the Riemann-Stieltjes integrability of  $f_i(\phi(t))$  with respect to  $x_i(t)$  implies the integrability of  $\int_C F \cdot d\vec{x}$ , the converse is not necessarily true.

**Exercise 13.25.** Compute the integral of 1-form on the three curves in Example 13.1.6. In case the three integrals are the same, can you provide a general reason?

1.  $x dx + y dy$ .
2.  $y dx - x dy$ .
3.  $(2x + ay)y dx + (x + by)x dy$ .
4.  $e^x(y dx + a dy)$ .

**Exercise 13.26.** Compute the integral of 1-form.

1.  $\int_C \frac{y dx - x dy}{x^2 + y^2}$ ,  $C$  is upper half circle in the counterclockwise direction.
2.  $\int_C (2a - y) dx + dy$ ,  $C$  is the cycloid  $x = at - b \sin t$ ,  $y = a - b \cos t$ ,  $0 \leq t \leq 2\pi$ .
3.  $\int_C x dy + y dz + z dx$ ,  $C$  is the straight line connecting  $(0, 0, 0)$  to  $(1, 1, 1)$ .
4.  $\int_C (x + z) dx + (y + z) dy + (x - y) dz$ ,  $C$  is the helix  $x = \cos t$ ,  $y = \sin t$ ,  $z = t$ ,  $0 \leq t \leq \alpha$ .
5.  $\int_C (y^2 - z^2) dx + (z^2 - x^2) dy + (x^2 - y^2) dz$ ,  $C$  is the boundary of the intersection of the sphere  $x^2 + y^2 + z^2 = 1$  with the first quadrant, and the direction is  $(x, y)$ -plane part, followed by  $(y, z)$ -plane part, followed by  $(z, x)$ -plane part.

**Exercise 13.27.** Suppose  $A$  is a symmetric matrix. Compute the integral of  $A\vec{x} \cdot d\vec{x}$  on any rectifiable curve.

**Exercise 13.28.** Prove that if  $F$  is continuous, then the 1-form  $F \cdot d\vec{x}$  is integrable along a rectifiable curve.

**Exercise 13.29.** Calculate the integral of the 1-form  $x dx + y dy$  along the graph  $(t, \kappa(t))$ ,  $t \in [0, 1]$ , of the Cantor function  $\kappa$  in Exercise 13.11.

**Exercise 13.30.** Prove that if the 1-form  $F \cdot d\vec{x}$  is integrable along a rectifiable curve and  $G$  is uniformly continuous on values of  $F$ , then  $(G \circ F) \cdot d\vec{x}$  is integrable.

**Exercise 13.31.** Prove that the sum and scalar multiplication of integrable 1-forms are integrable, and

$$\int_C (F + G) \cdot d\vec{x} = \int_C F \cdot d\vec{x} + \int_C G \cdot d\vec{x}, \quad \int_C cF \cdot d\vec{x} = c \int_C F \cdot d\vec{x}.$$

**Exercise 13.32.** Prove the inequality

$$\left| \int_C F \cdot d\vec{x} \right| \leq \mu(C) \sup_C \|F\|_2,$$

where  $\mu(C)$  is the Euclidean length of  $C$ .

**Exercise 13.33.** Suppose a rectifiable curve  $C$  is divided into two parts  $C_1$  and  $C_2$ . Prove that a 1-form is integrable on  $C$  if and only if it is integrable on  $C_1$  and  $C_2$ . Moreover,

$$\int_C F \cdot d\vec{x} = \int_{C_1} F \cdot d\vec{x} + \int_{C_2} F \cdot d\vec{x}.$$

**Exercise 13.34.** Suppose  $\vec{a}$  is a constant vector and  $U$  is an orthogonal linear transform. Prove that  $\int_{\vec{a}+U(C)} U(F(\vec{x})) \cdot d\vec{x} = \int_C F(\vec{a} + U(\vec{x})) \cdot d\vec{x}$ . Note that by Exercise 7.92, the transforms  $\vec{x} \mapsto \vec{a} + U(\vec{x})$  are exactly isometries of the Euclidean norm.

**Exercise 13.35.** Suppose  $c$  is a constant. Prove that  $\int_C F(\vec{x}) \cdot d\vec{x} = c \int_{cC} F(c\vec{x}) \cdot d\vec{x}$ .

**Exercise 13.36.** Define the oscillation  $\omega_C(F) = \sup_{\vec{x}, \vec{y} \in C} \|F(\vec{x}) - F(\vec{y})\|$ . Prove that if for any  $\epsilon > 0$ , there is  $\delta > 0$ , such that  $\|P\| < \delta$  implies  $\sum \omega_{C_i}(F)\mu(C_i) < \epsilon$ , where  $C_i$  are the segments of  $C$  cut by the partition  $P$ , then  $F \cdot d\vec{x}$  is integrable along  $C$ . Show that the converse is not true.

## 13.2 Surface

The length of a curve is defined by the approximation by straight line segments connecting partition points on the curve. For the area of a surface, the natural extension would be the approximation by triangles connecting points on the surface. However, the next example shows that the idea does not work.

**Example 13.2.1.** Consider the cylinder of radius 1 and height 1. Let  $\alpha = \frac{\pi}{m}$  and  $d = \frac{1}{2n}$ . At the heights  $0, d, 2d, 3d, \dots, 2nd$  of the cylinder, we have  $2n+1$  unit circles. On the unit circles at the even heights  $0, 2d, 4d, \dots, 2nd$ , we plot  $m$  points at angles  $0, 2\alpha, 4\alpha, \dots, (2m-2)\alpha$ . On the unit circles at the odd heights  $d, 3d, 5d, \dots, (2n-1)d$ , we plot  $m$  points at angles  $\alpha, 3\alpha, \dots, (2m-1)\alpha$ . By connecting nearby triple points, we get  $2mn$  identical isosceles triangles with base  $2 \sin \alpha$  and height  $\sqrt{(1 - \cos \alpha)^2 + d^2}$ . The total area is

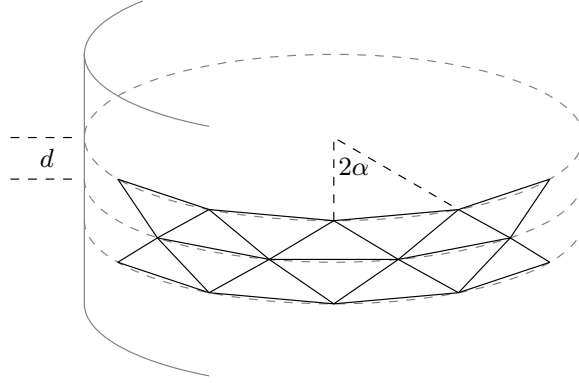
$$4mn \cdot \frac{1}{2} \cdot 2 \sin \alpha \cdot \sqrt{(1 - \cos \alpha)^2 + d^2} = 2\pi \frac{\sin \alpha}{\alpha} \sqrt{\left(\frac{1 - \cos \alpha}{d}\right)^2 + 1}.$$

This has no limit as  $\alpha, d \rightarrow 0$ .

### Area of Surface

The problem with the counterexample is that the directions of the approximate planes are not close to the directions of the tangent planes. So we need to use the “tangential approximation”, which for a curve is given by Exercise 13.13. In particular, we need to assume that the surface is continuously differentiable.





**Figure 13.2.1.** *A bad approximation for the area of cylinder.*

We take a surface to be a 2-dimensional submanifold (see Definition 8.4.1 in Section 8.4). Locally, such surfaces can either be described by a continuously differentiable regular parameterization (Proposition 8.4.2)

$$\sigma(u, v): U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^n,$$

or by the level of a continuously differentiable map at a regular value (Proposition 8.4.3). Here we take the parameterization viewpoint and assume  $S = \sigma(U)$ . Note that the assumption means that the surface is covered by one parameterization.

A partition  $Q$  of  $U$  is a decomposition of  $U$  into finitely many pieces  $I$ , such that the intersection between different pieces has zero area. The partition gives the corresponding partition  $P = \{\sigma(I)\}$  of  $S = \sigma(U)$ . For each piece  $I \in Q$ , choose a sample point  $\vec{u}_I^* \in I$ . Then the restriction  $\sigma|_I$  is approximated by the linear map

$$L_I(\vec{u}) = \sigma(\vec{u}_I^*) + \sigma'(\vec{u}_I^*)(\vec{u} - \vec{u}_I^*): \mathbb{R}^2 \rightarrow \mathbb{R}^n,$$

and the area of  $\sigma(I)$  is approximated by the area of  $L_I(I)$ .

What is the area (i.e., 2-dimensional measure) of  $L_I(I)$ ? The image of the linear transform  $L_I$  is, upon a choice of the origin, a 2-dimensional vector subspace of  $\mathbb{R}^n$ . With respect to the inner product on  $\mathbb{R}^n$ , the 2-dimensional vector subspace has a Lebesgue measure, defined as the unique translation invariant measure such that the measure  $\nu$  of the parallelogram spanned by  $\vec{x}$  and  $\vec{y}$  is the Euclidean length  $\|\vec{x} \times \vec{y}\|_2$  of the cross product (see (7.2.5)).

Now the composition  $\nu \circ L_I$  is a translation invariant measure on  $\mathbb{R}^2$ . By Theorem 11.4.4, we have  $\nu(L_I(X)) = c\mu(X)$  for all Lebesgue measurable  $X \subset \mathbb{R}^2$ , where  $c$  is a constant and  $\mu$  is the usual Lebesgue measure on  $\mathbb{R}^2$ . By taking  $X = [0, 1]^2$ , we get  $c = \nu(L_I([0, 1]^2))$ . Since  $L_I([0, 1]^2)$  is the parallelogram spanned by  $\sigma'(\vec{u}_I^*)(\vec{e}_1) = \sigma_u(\vec{u}_I^*)$  and  $\sigma'(\vec{u}_I^*)(\vec{e}_2) = \sigma_v(\vec{u}_I^*)$ , we get

$$c = \nu(L_I([0, 1]^2)) = \|\sigma_u(\vec{u}_I^*) \times \sigma_v(\vec{u}_I^*)\|_2,$$

and

$$\text{Area}(L_I(I)) = \nu(L_I(I)) = \|\sigma_u(\vec{u}_I^*) \times \sigma_v(\vec{u}_I^*)\|_2 \mu(I).$$

Thus the area of the surface is approximated by

$$\sum_{I \in Q} \text{Area}(L_I(I)) = \sum_{I \in Q} \|\sigma_u(\vec{u}_I^*) \times \sigma_v(\vec{u}_I^*)\|_2 \mu(I) = S(Q, \|\sigma_u \times \sigma_v\|_2).$$

This is the Riemann sum for the integral of the function  $\|\sigma_u \times \sigma_v\|_2$  on  $U$ . Therefore the surface area is

$$\mu(S) = \int_U \|\sigma_u \times \sigma_v\|_2 du dv = \int_U \sqrt{\|\sigma_u\|_2^2 \|\sigma_v\|_2^2 - (\sigma_u \cdot \sigma_v)^2} du dv.$$

We also denote

$$dA = \|\sigma_u \times \sigma_v\|_2 du dv,$$

where  $A$  indicates the area.

Intuitively, the area of a surface should be independent of the parameterization. To verify the intuition, we consider a continuously differentiable change of variable (the map is invertible, and both the map and its inverse are continuously differentiable)

$$(u, v) = (u(s, t), v(s, t)): V \rightarrow U.$$

We have  $\sigma_s = u_s \sigma_u + v_s \sigma_v$ ,  $\sigma_t = u_t \sigma_u + v_t \sigma_v$ , and by (7.2.4),

$$\sigma_s \times \sigma_t = \det \frac{\partial(u, v)}{\partial(s, t)} \sigma_u \times \sigma_v. \quad (13.2.1)$$

Then by the change of variable formula (Theorem 12.4.5), we have

$$\int_V \|\sigma_s \times \sigma_t\|_2 ds dt = \int_V \left| \det \frac{\partial(u, v)}{\partial(s, t)} \right| \|\sigma_u \times \sigma_v\|_2 ds dt = \int_U \|\sigma_u \times \sigma_v\|_2 du dv. \quad (13.2.2)$$

**Example 13.2.2.** The graph of a continuously differentiable function  $f(x, y)$  defined for  $(x, y) \in U$  is naturally parameterized as  $\sigma(x, y) = (x, y, f(x, y))$ . By

$$\sigma_x \times \sigma_y = (1, 0, f_x) \times (0, 1, f_y) = (-f_x, -f_y, 1),$$

the area of the surface is

$$\int_U \|(1, 0, f_x) \times (0, 1, f_y)\|_2 dx dy = \int_U \sqrt{1 + f_x^2 + f_y^2} dx dy.$$

In case  $z = f(x, y)$  is implicitly defined by  $g(x, y, z) = c$ , we may use the implicit differentiation to calculate  $z_x, z_y$  and then substitute as  $f_x, f_y$  above to get

$$\int_U \frac{\sqrt{g_x^2 + g_y^2 + g_z^2}}{|g_z|} dx dy = \int_U \frac{dx dy}{\cos \theta}.$$

Here  $\nabla g = (g_x, g_y, g_z)$  is normal to the tangent space, and  $\theta$  is the angle between the tangent plane and the  $(x, y)$ -plane.

**Example 13.2.3 (Pappus-Guldinus).** Suppose  $C$  is a continuously differentiable curve in  $\mathbb{R}^2$  parameterized by  $\phi(t) = (x(t), y(t))$ ,  $t \in [a, b]$ , such that  $y(t) > 0$ . The rotation of  $C$  with respect to the  $x$ -axis is the parameterized surface

$$\sigma(t, \theta) = (x(t), y(t) \cos \theta, y(t) \sin \theta): [a, b] \times [0, 2\pi) \rightarrow \mathbb{R}^3.$$

Since the tangent vectors  $\sigma_t = (x', y \cos \theta, y' \sin \theta)$  and  $\sigma_\theta = (0, -y \sin \theta, y \cos \theta)$  are orthogonal, we have

$$\|\sigma_t \times \sigma_\theta\|_2 = \|\sigma_t\|_2 \|\sigma_\theta\|_2 = y \sqrt{x'^2 + y'^2}.$$

Therefore the area is

$$\int_{[a,b] \times [0, 2\pi)} y \sqrt{x'^2 + y'^2} dt d\theta = 2\pi \int_a^b y \sqrt{x'^2 + y'^2} dt = 2\pi \int_C y ds.$$

In terms of the *center of weight* for the curve  $C$

$$(x_C^*, y_C^*) = \frac{1}{\mu(C)} \int_C (x, y) ds,$$

the area is  $2\pi y_C^* \mu(C)$ .

Note that  $y_C^*$  is the average of the distance from points in  $C$  to the  $x$ -axis. For a curve  $C$  lying on the positive side of a straight line  $L: ax + by = c$ , the area of the surface obtained by rotating  $C$  around  $L$  is  $2\pi d \mu(C)$ , where

$$d = \frac{ax_C^* + by_C^* - c}{\sqrt{a^2 + b^2}}$$

is the distance from the centre of weight  $(x_C^*, y_C^*)$  to  $L$ , and is also the average distance from the points on  $C$  to  $L$ .

For example, the torus is obtained by rotating a circle of radius  $a$  around a straight line of distance  $b$  away. The distance  $b$  is between the center of the circle and the line, and is also the average distance  $d$ . Therefore the torus has area  $4\pi^2 ab$ .

**Example 13.2.4.** The parameterized surface in  $\mathbb{R}^4$

$$\sigma(u, v) = (\cos(u+v), \sin(u+v), \cos u, \sin u), \quad u, v \in \left[0, \frac{\pi}{2}\right]$$

has area

$$\int_{[0, \frac{\pi}{2}]} \sqrt{\|\sigma_u\|_2^2 \|\sigma_v\|_2^2 - (\sigma_u \cdot \sigma_v)^2} du dv = \int_{[0, \frac{\pi}{2}]} \sqrt{2 \cdot 1 - 1^2} du dv = \frac{\pi^2}{4}.$$

**Exercise 13.37.** Find the area of the graph of a map  $(f(x, y), g(x, y)): \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , which is a surface in  $\mathbb{R}^4$ . More generally, find the area of the graph of a map  $F: \mathbb{R}^2 \rightarrow \mathbb{R}^{n-2}$ .

**Exercise 13.38.** Find the area of the surface in  $\mathbb{R}^4$  implicitly defined as  $f(x, y, z, w) = a$  and  $g(x, y, z, w) = b$ , similar to the formula in Example 13.2.2.

**Exercise 13.39.** Study the effect of transforms of  $\mathbb{R}^n$  on the area of surface.

1. For the shifting  $\vec{x} \rightarrow \vec{a} + \vec{x}$ , prove  $\mu(\vec{a} + S) = \mu(S)$ .
2. For the scaling  $\vec{x} \rightarrow c\vec{x}$ , prove  $\mu(cS) = c^2 \mu(S)$ .

3. For an orthogonal linear transform  $U$ , prove  $\mu(U(S)) = \mu(S)$ .

The discussion above assumes that the whole surface is covered by one (continuously differentiable and regular) parameterization. In general, we may not be able (or not convenient) to do so. Then we need to decompose  $S$  into finitely many smaller pieces  $S_i$ , so that the intersection among different pieces have zero area (for example, have dimension  $\leq 1$ ). The pieces should be small enough so that each is covered by one parameterization. Then we can use the earlier formula to compute the area of each  $S_i$ , and the sum of the areas of  $S_i$  is the area of  $S$ .

It remains to explain that the result of the computation is independent of the choice of the pieces  $S_i$  and their parameterizations. Suppose  $S'_j$  is another decomposition, and each  $S'_j$  has a regular parameterization. Then  $S$  is further decomposed into the intersections  $S''_{ij} = S_i \cap S'_j$ . Each  $S''_{ij}$  can be parameterized either as  $S_i$  or as  $S'_j$ . By  $S_i = \cup_j S''_{ij}$ , where the intersections of different pieces have zero area, the area of  $S$  computed from  $S_i$  is

$$\sum_i \mu(S_i) = \sum_i \sum_j \mu(S''_{ij}, \text{parameterized as } S_i).$$

Similarly, the area of  $S$  computed from  $S'_j$  is

$$\sum_j \mu(S'_j) = \sum_j \sum_i \mu(S''_{ij}, \text{parameterized as } S'_j).$$

Since we have argued in (13.2.2) that the two parameterizations of  $S''_{ij}$  give the same area  $\mu(S''_{ij})$ , the two sums are equal.

## Integration of Function Along Surface

The integration of a function  $f(\vec{x})$  along a surface  $S$  covered with one regular parameterization  $\sigma(u, v): U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^n$  is

$$\int_S f dA = \int_U f(\sigma(u, v)) \|\sigma_u(u, v) \times \sigma_v(u, v)\|_2 du dv.$$

In general, we may need to divide the surface into smaller parameterized pieces, such that the intersections of different pieces have zero area. Then we add the integrals on the pieces together.

By an argument similar to (13.2.2), the integral is independent of the choice of the parameterization

$$\begin{aligned} \int_S f dA &= \int_V f(\sigma(s, t)) \|\sigma_s \times \sigma_t\|_2 ds dt \\ &= \int_V f(\sigma(s, t)) \left| \det \frac{\partial(u, v)}{\partial(s, t)} \right| \|\sigma_u \times \sigma_v\|_2 ds dt \\ &= \int_U f(\sigma(u, v)) \|\sigma_u \times \sigma_v\|_2 du dv. \end{aligned} \tag{13.2.3}$$

Here the first equality is the definition of  $\int_S f dA$  in terms of the parameter  $(s, t)$ , the second is the change of variable formula, the third is by (13.2.1), and the last expression is the definition of  $\int_S f dA$  in terms of the parameter  $(u, v)$ .

We may further show that the integral is also independent of the decomposition of the surface into parameterized pieces. Moreover, the integral has properties similar to the integral of functions along rectifiable curves.

**Example 13.2.5.** For a fixed vector  $\vec{a} \in \mathbb{R}^3$ , consider the integral  $\int_{S^2} f(\vec{a} \cdot \vec{x}) dA$  on the unit sphere. There is a rotation that moves  $\vec{a}$  to  $(r, 0, 0)$ ,  $r = \|\vec{a}\|_2$ . Since the rotation preserves the surface area, we have

$$\int_{S^2} f(\vec{a} \cdot \vec{x}) dA = \int_{S^2} f((r, 0, 0) \cdot (x, y, z)) dA = \int_{S^2} f(rx) dA.$$

Parametrize  $S^2$  by  $\sigma(x, \theta) = (x, \sqrt{1-x^2} \cos \theta, \sqrt{1-x^2} \sin \theta)$ . Then  $\|\sigma_x \times \sigma_\theta\|_2 = 1$  and

$$\int_{S^2} f(rx) dA = \int_{\substack{-1 \leq x \leq 1 \\ 0 \leq \theta \leq 2\pi}} f(rx) dx d\theta = 2\pi \int_{-1}^1 f(rx) dx = \frac{2\pi}{\|\vec{a}\|_2} \int_{-\|\vec{a}\|_2}^{\|\vec{a}\|_2} f(x) dx.$$

**Exercise 13.40.** Study the effect of transforms of  $\mathbb{R}^n$  on the integral along surface.

1. For the shifting  $\vec{x} \rightarrow \vec{a} + \vec{x}$ , prove  $\int_{\vec{a}+S} f(\vec{x}) dA = \int_S f(\vec{a} + \vec{x}) dA$ .
2. For the scaling  $\vec{x} \rightarrow c\vec{x}$ , prove  $\int_{cS} f(\vec{x}) dA = c^2 \int_S f(c\vec{x}) dA$ .
3. For an orthogonal linear transform  $U$ , prove  $\int_{U(S)} f(\vec{x}) dA = \int_S f(U(\vec{x})) dA$ .

**Exercise 13.41.** Compute the attracting force  $\int_{S^2} \frac{\vec{x} - \vec{a}}{\|\vec{x} - \vec{a}\|_2^3} dA$  of the unit sphere on a point  $\vec{a}$ .

**Exercise 13.42.** Prove that the definition of  $\int_S f dA$  is independent of the decomposition of the surface into parameterized pieces.

## Integration of 2-Form Along Surface

Suppose  $\sigma(u, v): U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$  is a continuously differentiable regular parameterization of a surface  $S \subset \mathbb{R}^3$ . Then the parameterization gives a *normal vector*

$$\vec{n} = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|_2}.$$

Geometrically, the normal vector is determined by the following properties:

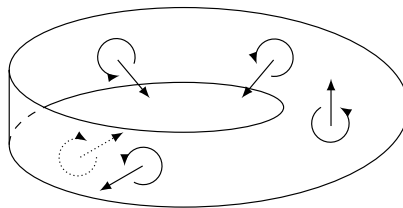
- Direction:  $\vec{n}$  is orthogonal to the tangent plane of  $S$ .

- Length:  $\vec{n}$  has Euclidean norm 1.
- Orientation:  $\vec{n}, \sigma_u, \sigma_v$  follows the “right hand rule”, which means that  $\det(\vec{n} \ \sigma_u \ \sigma_v) > 0$ .

The first property means  $\vec{n} \cdot \sigma_u = 0$  and  $\vec{n} \cdot \sigma_v = 0$ . The first two properties are independent of the choice of the parameterization and determine the normal vector up to the choice among  $\vec{n}$  and  $-\vec{n}$ . The orientation picks one from the two choices.

In general, we may need to divide the surface into parameterized pieces  $S_i$ . Then we need the normal vectors of the piece to be compatible in the sense that the normal vector changes continuously when we move from one piece to another. Then the whole surface has a *continuous choice* of the normal vectors. Such a choice is called an *orientation*, and the surface is called *orientable*.

It is possible for a surface to be *nonorientable* in the sense that there is no continuous choice of the normal vectors. This is equivalent to the nonexistence of a collection of compatibly oriented parameterizations. A typical example is the Möbius band. In fact, a surface is not orientable if and only if it contains a Möbius band.



**Figure 13.2.2.** The Möbius band is not orientable.

**Example 13.2.6.** The graph of a continuously differentiable function  $f(x, y)$  is naturally parameterized as  $\sigma(x, y) = (x, y, f(x, y))$ . By

$$\sigma_x \times \sigma_y = (1, 0, f_x) \times (0, 1, f_y) = (-f_x, -f_y, 1),$$

the normal vector induced by the parameterization is

$$\vec{n} = \frac{\sigma_x \times \sigma_y}{\|\sigma_x \times \sigma_y\|} = \frac{(-f_x, -f_y, 1)}{\sqrt{f_x^2 + f_y^2 + 1}}.$$

Alternatively, we may try to find  $\vec{n}$  by using the three characteristic properties. First we solve for  $\vec{v} = (a, b, c)$

$$\vec{v} \cdot \sigma_x = a + cf_x = 0, \quad \vec{v} \cdot \sigma_y = b + cf_y = 0.$$

Choosing  $c = 1$ , we get  $\vec{v} = (-f_x, -f_y, 1)$ , which is parallel to  $\vec{n}$ . Then by

$$\det(\vec{v} \ \sigma_x \ \sigma_y) = \det \begin{pmatrix} -f_x & 1 & 0 \\ -f_y & 0 & 1 \\ 1 & f_x & f_y \end{pmatrix} = f_x^2 + f_y^2 + 1 > 0,$$

we know  $\vec{n}$  and  $\vec{v}$  point to the same direction. Since  $\vec{n}$  has the unit length, we get

$$\vec{n} = \frac{\vec{v}}{\|\vec{v}\|} = \frac{(-f_x, -f_y, 1)}{\sqrt{f_x^2 + f_y^2 + 1}}.$$

**Example 13.2.7.** The sphere  $S_R^2 = \{(x, y, z) : x^2 + y^2 + z^2 = R^2\}$  of radius  $R$  can be covered by the parameterizations

$$\begin{aligned}\sigma_1(x, y) &= (x, y, \sqrt{R^2 - x^2 - y^2}), & x^2 + y^2 < R^2, \\ \sigma_2(y, x) &= (x, y, -\sqrt{R^2 - x^2 - y^2}), & x^2 + y^2 < R^2, \\ \sigma_3(z, x) &= (x, \sqrt{R^2 - x^2 - z^2}, z), & x^2 + z^2 < R^2, \\ \sigma_4(x, z) &= (x, -\sqrt{R^2 - x^2 - z^2}, z), & x^2 + z^2 < R^2, \\ \sigma_5(y, z) &= (\sqrt{R^2 - y^2 - z^2}, y, z), & y^2 + z^2 < R^2, \\ \sigma_6(z, y) &= (-\sqrt{R^2 - y^2 - z^2}, y, z), & y^2 + z^2 < R^2.\end{aligned}$$

By

$$(\sigma_1)_x \times (\sigma_1)_y = \left(1, 0, -\frac{x}{z}\right) \times \left(0, 1, -\frac{y}{z}\right) = \left(\frac{x}{z}, \frac{y}{z}, 1\right),$$

the normal vector from the first parameterization is

$$\frac{(\sigma_1)_x \times (\sigma_1)_y}{\|(\sigma_1)_x \times (\sigma_1)_y\|_2} = \frac{(x, y, z)}{\|(x, y, z)\|_2} = \frac{\vec{x}}{R}.$$

Similar computations also show that  $\frac{\vec{x}}{R}$  is the normal vector from the other parameterizations. Therefore the choice  $\vec{n} = \frac{\vec{x}}{R}$  gives an orientation of the sphere.

Note that the order of the variables in  $\sigma_2$  are deliberately arranged to have  $y$  as the first and  $x$  as the second. Thus the normal vector should be computed from  $(\sigma_2)_y \times (\sigma_2)_x$  instead of the other way around.

Like a vector field along a curve, a vector field along a surface is a map  $F: S \rightarrow \mathbb{R}^n$ . If  $S$  is orientable, with orientation  $\vec{n}$ , then the *flux* of the vector field along the surface is

$$\int_S F \cdot \vec{n} dA = \int_U F(\sigma(u, v)) \cdot (\sigma_u(u, v) \times \sigma_v(u, v)) du dv.$$

The left side shows that the integral is independent of the choice of the parameterization, as long as the parameterization is *compatible* with the orientation, in the sense that  $\sigma_u \times \sigma_v$  is a positive multiple of  $\vec{n}$ . In fact, changing the orientation from  $\vec{n}$  to  $-\vec{n}$  changes the integral by a negative sign. This is similar to the effect on  $\int_C F \cdot d\vec{x}$  when we reverse the orientation of  $C$ .

**Example 13.2.8.** The outgoing flux of the flow  $F = (x^2, y^2, z^2)$  through the sphere  $\|\vec{x} - \vec{x}_0\|_2^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = R^2$  is

$$\int_{\|\vec{x} - \vec{x}_0\|_2 = R} (x^2, y^2, z^2) \cdot \frac{(x - x_0, y - y_0, z - z_0)}{R} dA.$$

By the shifting  $\vec{x} \rightarrow \vec{x}_0 + \vec{x}$ , the integral is equal to

$$\frac{1}{R} \int_{\|\vec{x}\|_2 = R} ((x + x_0)^2 x + (y + y_0)^2 y + (z + z_0)^2 z) dA.$$

By suitable rotations that exchange the axis, we have

$$\int_{\|\vec{x}\|_2=R} x^n dA = \int_{\|\vec{x}\|_2=R} y^n dA = \int_{\|\vec{x}\|_2=R} z^n dA.$$

By the transform  $(x, y, z) \rightarrow (-x, y, z)$ , we also know that the first integral vanishes for odd integers  $n$ . Thus the integral above becomes

$$\frac{1}{R} \int_{\|\vec{x}\|_2=R} (2x_0x^2 + 2y_0y^2 + 2z_0z^2) dA = \frac{2}{R} (x_0 + y_0 + z_0) \int_{\|\vec{x}\|_2=R} x^2 dA.$$

Then we further use the equalities above to get

$$\int_{\|\vec{x}\|_2=R} x^2 dA = \frac{1}{3} \int_{\|\vec{x}\|_2=R} (x^2 + y^2 + z^2) dA = \frac{1}{3} \int_{\|\vec{x}\|_2=R} R^2 dA = \frac{1}{3} R^2 \cdot 4\pi R^2 = \frac{4}{3} \pi R^4.$$

We conclude that the flux is  $\frac{8}{3} \pi (x_0 + y_0 + z_0) R^3$ .

**Exercise 13.43.** Study the effect of transforms of  $\mathbb{R}^n$  on the flux, similar to Exercise 13.40.

**Exercise 13.44.** By an argument similar to (13.2.3), prove that  $\int_S F \cdot \vec{n} dA$  is not changed by compatible change of variable.

Now we give a formal interpretation of the flux. For the parameterization  $\sigma(u, v) = (x(u, v), y(u, v), z(u, v))$ , by formally taking the cross product of differential 1-forms, we have

$$\begin{aligned} dx \times dy &= (x_u du + x_v dv) \times (y_u du + y_v dv) \\ &= (x_u y_v - x_v y_u) du \times dv = \det \frac{\partial(x, y)}{\partial(u, v)} du \times dv, \end{aligned}$$

and the similar formulae for  $dy \times dz$  and  $dz \times dx$ . Then for  $F = (f, g, h)$ , we have

$$\begin{aligned} [F \cdot (\sigma_u \times \sigma_v)] du \times dv &= [f(y_u z_v - y_v z_u) + g(z_u x_v - z_v x_u) + h(x_u y_v - x_v y_u)] du \times dv \\ &= f dy \times dz + g dz \times dx + h dx \times dy. \end{aligned}$$

This suggests that the flux  $\int_S F \cdot \vec{n} dA$  should really be denoted  $\int_S f dy \times dz + g dz \times dx + h dx \times dy$ .

Since the cross product is a special case of the exterior product, we introduce the *differential 2-form* on  $\mathbb{R}^3$

$$\omega = f dy \wedge dz + g dz \wedge dx + h dx \wedge dy.$$

Then the flux may be interpreted as *the integral of the differential 2-form* on an oriented surface

$$\begin{aligned} &\int_S f dy \wedge dz + g dz \wedge dx + h dx \wedge dy \\ &= \int_U \left( f \det \frac{\partial(y, z)}{\partial(u, v)} + g \det \frac{\partial(z, x)}{\partial(u, v)} + h \det \frac{\partial(x, y)}{\partial(u, v)} \right) du dv. \end{aligned} \quad (13.2.4)$$



The formula is only for one piece of the regular parameterization  $\sigma(u, v)$ , such that  $\sigma_u \times \sigma_v$  is a positive multiple of  $\vec{n}$ . In general, we need to divide the surface into such pieces and add the integrals on the pieces together.

The use of exterior product allows us to extend the integral of 2-forms to more general orientable surfaces. A *differential 2-form* on  $\mathbb{R}^n$  is

$$\omega = \sum_{1 \leq i < j \leq n} f_{ij}(\vec{x}) dx_i \wedge dx_j,$$

and its integral on a surface  $S \subset \mathbb{R}^n$  regularly parameterized by continuously differentiable map  $\sigma(u, v): U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^n$  is

$$\int_S \omega = \sum_{1 \leq i < j \leq n} \int_U f_{ij}(\sigma(u, v)) \det \frac{\partial(x_i, x_j)}{\partial(u, v)} du dv. \quad (13.2.5)$$

Under a change of variable between  $(u, v) \in U$  and  $(s, t) \in V$ , we have (see Theorem 12.4.5)

$$\begin{aligned} & \int_U f(\sigma(u, v)) \det \frac{\partial(x_i, x_j)}{\partial(u, v)} du dv \\ &= \int_V f(\sigma(s, t)) \det \frac{\partial(x_i, x_j)}{\partial(u, v)} \left| \det \frac{\partial(u, v)}{\partial(s, t)} \right| ds dt. \end{aligned}$$

Since  $\frac{\partial(x_i, x_j)}{\partial(s, t)} = \frac{\partial(x_i, x_j)}{\partial(u, v)} \frac{\partial(u, v)}{\partial(s, t)}$ , the right side fits into the definition (13.2.6) in terms of the new variable  $(s, t)$  if and only if  $\det \frac{\partial(u, v)}{\partial(s, t)} > 0$ . This is the condition for (13.2.6) to be well defined.

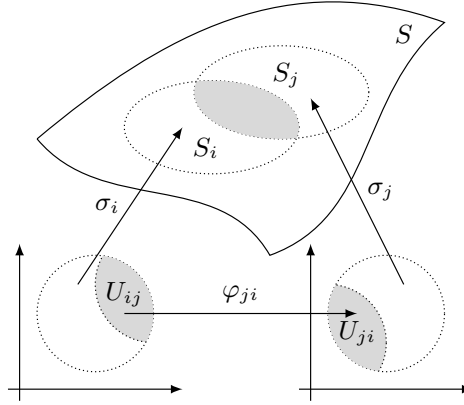
In case  $n = 3$ , by (13.2.1), the condition  $\det \frac{\partial(u, v)}{\partial(s, t)} > 0$  means exactly that both parameterizations give the same normal vector. Although we no longer have the normal vector for general  $n$ , we may still talk about compatible parameterizations in the sense that  $\det \frac{\partial(u, v)}{\partial(s, t)} > 0$ . Therefore we say a surface  $S \subset \mathbb{R}^n$  is *oriented* if it is covered by finitely many parameterized pieces  $\sigma_i: U_i \subset \mathbb{R}^2 \rightarrow S_i \subset \mathbb{R}^n$ , where  $S_i$  are *open* subsets of  $S$  and  $S = \cup S_i$ , such that the *transition maps* (see Figure 14.1.1)

$$\varphi_{ji} = \sigma_j^{-1} \sigma_i: U_{ij} = \sigma_i^{-1}(S_i \cap S_j) \rightarrow U_{ji} = \sigma_j^{-1}(S_i \cap S_j)$$

on the overlapping of pieces satisfy

$$\det \varphi'_{ji} > 0 \text{ for any } i, j.$$

Such a collection  $\{\sigma_i\}$  gives an *orientation* of the surface. Any other parameterization  $\sigma_*: U \subset \mathbb{R}^2 \rightarrow S_* \subset S$  is compatible with the orientation if  $\det(\sigma_i^{-1} \circ \sigma_*)' > 0$  on  $\sigma^{-1}(S_* \cap S_i)$ .



**Figure 13.2.3.** Transition map between overlapping parameterizations.

We would like to define the integral of a 2-form  $\omega$  on an oriented surface as

$$\int_S \omega = \sum_k \int_{S_k} \omega = \sum_k \sum_{1 \leq i < j \leq n} \int_{U_k} f_{ij}(\sigma_k(u, v)) \det \frac{\partial(x_i, x_j)}{\partial(u, v)} du dv.$$

However, we need to take into account of the overlapping between pieces, where the integral is computed more than once. A practical way to get around this is to find subsets  $B_k \subset U_k$ , such that  $S = \cup_k \sigma_k(B_k)$  and  $\sigma_k(B_k) \cap \sigma_l(B_l)$  has zero area for any  $k \neq l$ . Then define

$$\int_S \omega = \sum_k \sum_{1 \leq i < j \leq n} \int_{B_k} f_{ij}(\sigma_k(u, v)) \det \frac{\partial(x_i, x_j)}{\partial(u, v)} du dv. \quad (13.2.6)$$

A more theoretical way is to use the *partition of unity*. See the theory of calculus on manifolds.

**Example 13.2.9.** We compute the outgoing flux of  $F = (x^2, y^2, z^2)$  in Example 13.2.8 through the ellipse  $S$  given by  $\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} + \frac{(z-z_0)^2}{c^2} = 1$ . By the shifting  $\vec{x} \rightarrow \vec{x}_0 + \vec{x}$ , the flux is the integral

$$\int_{S+\vec{x}_0} (x+x_0)^2 dy \wedge dz + (y+y_0)^2 dz \wedge dx + (z+z_0)^2 dx \wedge dy.$$

The integral of  $(z+z_0)^2 dx \wedge dy$  may be computed by using parameterizations similar to

$\sigma_1$  and  $\sigma_2$  in Example 13.2.7

$$\begin{aligned}
 \int_{S-\vec{x}_0} (z+z_0)^2 dx \wedge dy &= \int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1} \left( c\sqrt{1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}} + z_0 \right)^2 \det \frac{\partial(x,y)}{\partial(x,y)} dx dy \\
 &\quad + \int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1} \left( -c\sqrt{1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}} + z_0 \right)^2 \det \frac{\partial(x,y)}{\partial(y,x)} dx dy \\
 &= 4z_0 c \int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1} \sqrt{1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}} dx dy \\
 &= 4abc z_0 \int_{u^2+v^2 \leq 1} \sqrt{1-u^2-v^2} du dv = \frac{8}{3} \pi abc z_0.
 \end{aligned}$$

The total flux is  $\frac{8}{3} \pi abc(x_0 + y_0 + z_0)$ .

**Example 13.2.10.** Let  $S$  be the surface parameterised by  $\sigma(u, v) = (u+v, u^2+v^2, \dots, u^n+v^n)$  for  $0 \leq u \leq a$ ,  $0 \leq v \leq b$ . The parameterization gives an orientation of the surface. With respect to this orientation, the integral

$$\begin{aligned}
 &\int_S dx_1 \wedge dx_2 + dx_2 \wedge dx_3 + \dots + dx_{n-1} \wedge dx_n \\
 &= \int_S \left( \det \frac{\partial(x_1, x_2)}{\partial(u, v)} + \det \frac{\partial(x_2, x_3)}{\partial(u, v)} + \dots + \det \frac{\partial(x_{n-1}, x_n)}{\partial(u, v)} \right) du dv \\
 &= \int_0^a \int_0^b \left[ \det \begin{pmatrix} 1 & 1 \\ 2u & 2v \end{pmatrix} + \det \begin{pmatrix} 2u & 2v \\ 3u^2 & 3v^2 \end{pmatrix} \right. \\
 &\quad \left. + \dots + \det \begin{pmatrix} (n-1)u^{n-2} & (n-1)v^{n-2} \\ nu^{n-1} & nv^{n-1} \end{pmatrix} \right] du dv \\
 &= \int_0^a \int_0^b [1 \cdot 2(v-u) + 2 \cdot 3(uv^2 - u^2v) + \dots + (n-1)n(u^{n-2}v^{n-1} - u^{n-1}v^{n-2})] du dv \\
 &= (ab^2 - a^2b) + (a^2b^3 - a^3b^2) + \dots + (a^{n-1}b^n - a^n b^{n-1}) \\
 &= (b-a)(ab + a^2b^2 + \dots + a^{n-1}b^{n-1}) = \frac{(b-a)(1-a^n b^n)}{1-ab}.
 \end{aligned}$$

**Exercise 13.45.** Compute the integral of 2-form.

1.  $\int_S xy^2 dy \wedge dz + yz^2 dz \wedge dx + zx^2 dx \wedge dy$ ,  $S$  is the ellipse  $\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} + \frac{(z-z_0)^2}{c^2} = 1$  with orientation given by the inward normal vector.
2.  $\int_S xyz dx \wedge dy$ ,  $S$  is the part of the ellipse  $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ ,  $x \geq 0$ ,  $y \geq 0$ , with orientation given by outward normal vector.
3.  $\int_S xy dy \wedge dz$  and  $\int_S x^2 y dy \wedge dz$ ,  $S$  is the boundary of the solid enclosed by  $z = x^2 + y^2$  and  $z = 1$ , with orientation given by outward normal vector.

**Exercise 13.46.** Prove that the area of a surface  $S \subset \mathbb{R}^3$  given by an equation  $g(x, y, z) = c$  is  $\int_S \frac{g_x dy \wedge dz + g_y dz \wedge dx + g_z dx \wedge dy}{\sqrt{g_x^2 + g_y^2 + g_z^2}}$ .

### 13.3 Submanifold

The integration along curves and surfaces can be extended to higher dimensional submanifolds.

#### Volume and Integration of Function

Suppose  $M$  is a  $k$ -dimensional submanifold of  $\mathbb{R}^n$ . If  $M$  is regularly parameterized by continuously differentiable  $\sigma(\vec{u}): U \subset \mathbb{R}^k \rightarrow \mathbb{R}^n$ , then the volume of  $M$  is

$$\mu(M) = \int_U \|\sigma_{u_1} \wedge \cdots \wedge \sigma_{u_k}\|_2 du_1 \cdots du_k.$$

The definition is independent of the choice of parameterization. We also denote the volume unit

$$\begin{aligned} dV &= \|\sigma_{u_1} \wedge \cdots \wedge \sigma_{u_k}\|_2 du_1 \cdots du_k \\ &= \sqrt{\det(\sigma_{u_i} \cdot \sigma_{u_j})_{1 \leq i, j \leq k}} du_1 \cdots du_k \\ &= \sqrt{\sum_{1 \leq i_1 < \cdots < i_n \leq N} \left( \det \frac{\partial(x_{i_1}, \dots, x_{i_n})}{\partial(u_1, \dots, u_n)} \right)^2} du_1 \cdots du_n. \end{aligned} \quad (13.3.1)$$

The integral of a function along the submanifold is

$$\int_M f dV = \int_U f(\sigma(u_1, \dots, u_k)) \|\sigma_{u_1} \wedge \cdots \wedge \sigma_{u_k}\|_2 du_1 \cdots du_k$$

The integral is also independent of the choice of parameterization.

**Example 13.3.1.** Let  $F(\vec{\theta}): U \subset \mathbb{R}^{n-1} \rightarrow \mathbb{R}^n$  be a parameterization of the unit sphere  $S^{n-1}$ . Since  $F \cdot F = \|F\|_2^2 = 1$ , by taking partial derivatives, we get  $F_{\theta_i} \cdot F = 0$ . Therefore  $F$  is a unit length vector orthogonal to  $F_{\theta_i}$ . This implies that

$$\|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 = \|F \wedge F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 = |\det(F \ F')|,$$

so that the volume of the sphere  $S^{n-1}$  is

$$\beta_{n-1} = \int_U \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 d\mu_{\vec{\theta}} = \int_U |\det(F \ F')| d\mu_{\vec{\theta}}.$$

The parameterization  $F(\vec{\theta})$  induces a parameterization of  $S^n$

$$G(\vec{\theta}, \phi) = (F(\vec{\theta}) \cos \phi, \sin \phi): U \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \rightarrow \mathbb{R}^{n+1}.$$

We have

$$\begin{aligned}
& \|G_{\theta_1} \wedge G_{\theta_2} \wedge \cdots \wedge G_{\theta_{n-1}} \wedge G_\phi\|_2 \\
&= \|(F_{\theta_1} \cos \phi, 0) \wedge (F_{\theta_2} \cos \phi, 0) \wedge \cdots \wedge (F_{\theta_{n-1}} \cos \phi, 0) \wedge (-F \sin \phi, \cos \phi)\|_2 \\
&= \|(F_{\theta_1} \cos \phi, 0) \wedge (F_{\theta_2} \cos \phi, 0) \wedge \cdots \wedge (F_{\theta_{n-1}} \cos \phi, 0)\|_2 \|(-F \sin \phi, \cos \phi)\|_2 \\
&= \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 \cos^{n-1} \phi,
\end{aligned}$$

where the second equality is due to  $(-F \sin \phi, \cos \phi) \perp (F_{\theta_i} \cos \phi, 0)$ , and the last equality is due to

$$\|(-F \sin \phi, \cos \phi)\|_2 = \sqrt{\|F\|_2^2 \sin^2 \phi + \cos^2 \phi} = \sqrt{\sin^2 \phi + \cos^2 \phi} = 1.$$

Therefore

$$\beta_n = \int_{-\frac{\pi}{2} \leq \phi \leq \frac{\pi}{2}}^{\bar{\theta} \in A} \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 \cos^{n-1} \phi d\mu_{\bar{\theta}} d\phi = \beta_{n-1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{n-1} \phi d\phi.$$

Using integration by parts, we get

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^n \phi d\phi = \frac{n-1}{n} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{n-2} \phi d\phi = \cdots = \begin{cases} \frac{(n-1)(n-3) \cdots 1}{n(n-2) \cdots 2} \pi, & \text{if } n \text{ is even,} \\ \frac{(n-1)(n-3) \cdots 2}{n(n-2) \cdots 3} 2, & \text{if } n \text{ is odd.} \end{cases}$$

Therefore  $\beta_n = \frac{2\pi}{n-1} \beta_{n-2}$ . Combined with  $\beta_1 = 2\pi$ ,  $\beta_2 = 4\pi$ , we get

$$\beta_{n-1} = \begin{cases} \frac{2\pi^k}{(k-1)!}, & \text{if } n = 2k, \\ \frac{2^{k+1}\pi^k}{(2k-1)(2k-3) \cdots 1}, & \text{if } n = 2k+1. \end{cases}$$

We also note that

$$H(r, \vec{\theta}) = rF(\vec{\theta}): (0, \infty) \times A \rightarrow \mathbb{R}^n$$

is a *spherical change of variable*. We have  $H' = (F \ rF')$ ,  $\det H' = r^{n-1} \det(F \ F')$ , so that

$$\int_{a \leq \|\vec{x}\|_2 \leq b} f(\|\vec{x}\|_2) d\mu = \int_{\substack{\bar{\theta} \in A \\ a \leq r \leq b}} f(r) r^{n-1} |\det(F \ F')| dr d\mu_{\bar{\theta}} = \beta_{n-1} \int_a^b f(r) r^{n-1} dr.$$

In particular, by taking  $f = 1$ ,  $a = 0$ ,  $b = 1$ , we get the volume of the unit ball  $B^n$

$$\alpha_n = \beta_{n-1} \int_0^1 r^{n-1} dr = \frac{\beta_{n-1}}{n} = \begin{cases} \frac{\pi^k}{k!}, & \text{if } n = 2k, \\ \frac{2^{k+1}\pi^k}{(2k+1)(2k-1)(2k-3) \cdots 1}, & \text{if } n = 2k+1. \end{cases}$$

**Example 13.3.2.** Consider a  $k$ -dimensional submanifold  $M$  parameterized by

$$\sigma(\vec{u}) = (\xi(\vec{u}), r(\vec{u})): U \subset \mathbb{R}^k \rightarrow \mathbb{R}^n \times \mathbb{R}.$$

Assume  $r(\vec{u}) \geq 0$ , so that the submanifold is inside the upper half Euclidean space. The  $l$ -dimensional rotation of  $M$  around the axis  $\mathbb{R}^n \times 0$  is a  $(k+l)$ -dimensional submanifold

$$\rho_l(M) = \{(\vec{x}, \vec{y}) \in \mathbb{R}^n \times \mathbb{R}^{l+1} : (\vec{x}, \|\vec{y}\|_2) \in S\}.$$

Let  $F(\vec{\theta}) : V \subset \mathbb{R}^l \rightarrow \mathbb{R}^{l+1}$  be a parameterization of the unit sphere  $S^l$ . Then the rotation submanifold  $\rho_l(M)$  is parameterized by

$$\rho(\vec{u}, \vec{\theta}) = (\xi(\vec{u}), r(\vec{u})F(\vec{\theta})) : U \times V \rightarrow \mathbb{R}^n \times \mathbb{R}^{l+1}.$$

By  $\rho_{u_i} = (\xi_{u_i}, r_{u_i}F)$ ,  $\rho_{\theta_j} = (\vec{0}, rF_{\theta_j})$ , and  $\rho_{u_i} \cdot \rho_{\theta_j} = r_{u_i}F \cdot rF_{\theta_j} = 0$ , we get

$$\|\rho_{u_1} \wedge \rho_{u_2} \wedge \cdots \wedge \rho_{u_k} \wedge \rho_{\theta_1} \wedge \rho_{\theta_2} \wedge \cdots \wedge \rho_{\theta_l}\|_2 = \|\rho_{u_1} \wedge \rho_{u_2} \wedge \cdots \wedge \rho_{u_k}\|_2 \|\rho_{\theta_1} \wedge \rho_{\theta_2} \wedge \cdots \wedge \rho_{\theta_l}\|_2,$$

and

$$\|\rho_{\theta_1} \wedge \rho_{\theta_2} \wedge \cdots \wedge \rho_{\theta_l}\|_2 = r^l \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_l}\|_2.$$

Moreover, there is an orthogonal transform  $U_F$  on  $\mathbb{R}^{l+1}$  such that  $U_F(F) = (1, 0, \dots, 0)$ . Then  $(\text{id}, U_F)$  is an orthogonal transform on  $\mathbb{R}^n \times \mathbb{R}^{l+1}$  such that

$$(\text{id}, U_F)\rho_{u_i} = (\xi_{u_i}, r_{u_i}U_F(F)) = (\xi_{u_i}, r_{u_i}, 0, \dots, 0) = (\sigma_{u_i}, 0, \dots, 0).$$

Applying the orthogonal transform to  $\|\rho_{u_1} \wedge \rho_{u_2} \wedge \cdots \wedge \rho_{u_k}\|_2$ , we get

$$\|\rho_{u_1} \wedge \rho_{u_2} \wedge \cdots \wedge \rho_{u_k}\|_2 = \|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_k}\|_2$$

Thus the volume of the rotation hypersurface is

$$\begin{aligned} \mu(\rho_l(S)) &= \int_{U \times V} \|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_k}\|_2 r^l \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_l}\|_2 d\mu_{\vec{u}} d\mu_{\vec{\theta}} \\ &= \beta_l \int_U \|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_k}\|_2 r^l d\mu_{\vec{u}} = \beta_l \int_S r^l dV. \end{aligned}$$

**Exercise 13.47.** Use the rotation in Example 13.3.2 to prove that

$$\beta_{k+l+1} = \beta_k \beta_l \int_0^{\frac{\pi}{2}} \cos^k \theta \sin^l \theta d\theta.$$

**Exercise 13.48.** Extend the result of Example 13.3.2 to the rotation around a hyperplane  $\vec{a} \cdot \vec{x} = b$  (the surface is on the positive side of the hyperplane).

## Integration of Differential Form

A *differential  $k$ -form* on  $\mathbb{R}^n$  is

$$\omega = \sum_{i_1 < \cdots < i_k} f_{i_1 \cdots i_k}(\vec{x}) dx_{i_1} \wedge \cdots \wedge dx_{i_k}. \quad (13.3.2)$$

Its integral on a regularly parameterized  $k$ -dimensional submanifold  $M$  is

$$\int_M \omega = \int_U f_{i_1 \cdots i_k}(\sigma(u_1, \dots, u_k)) \det \frac{\partial(x_{i_1}, \dots, x_{i_k})}{\partial(u_1, \dots, u_k)} du_1 \cdots du_k.$$

Similar to the integral of differential 2-forms on surfaces, the integral is independent of the parameterizations as long as they are compatibly oriented. The compatibility means that the determinant of the Jacobian for the change of variables (so called transition map) is positive. In general, the submanifold may be divided into several compatibly oriented pieces, and the integral of the differential form is the sum of the integrals on the pieces (strictly speaking, on smaller pieces such that the overlappings have zero volume).

Now we consider the special case of the integral of an  $(n-1)$ -form on an oriented  $(n-1)$ -dimensional submanifold  $M \subset \mathbb{R}^n$ . What is special here is the existence of the normal vector. Similar to parameterized surfaces in  $\mathbb{R}^3$ , the normal vector is determined by the following properties:

- Direction:  $\vec{n} \cdot \sigma_{u_i} = 0$  for  $i = 1, \dots, n-1$ .
- Length:  $\|\vec{n}\|_2 = 1$ .
- Orientation:  $\det(\vec{n} \sigma_{u_1} \cdots \sigma_{u_{n-1}}) > 0$ .

By the first property and (7.4.2), we have

$$(\sigma_{u_1} \wedge \cdots \wedge \sigma_{u_{n-1}})^* = c\vec{n},$$

where

$$c = \det(\sigma_{u_1} \cdots \sigma_{u_{n-1}} \vec{n}) = (-1)^{n-1} \det(\vec{n} \sigma_{u_1} \cdots \sigma_{u_{n-1}}).$$

Then by the second and third properties, we get

$$\vec{n} = (-1)^{n-1} \frac{(\sigma_{u_1} \wedge \cdots \wedge \sigma_{u_{n-1}})^*}{\|\sigma_{u_1} \wedge \cdots \wedge \sigma_{u_{n-1}}\|_2}.$$

Let ( $\widehat{x_i}$  means the term  $x_i$  is missing from the list)

$$\begin{aligned} \vec{a} &= (\sigma_{u_1} \wedge \cdots \wedge \sigma_{u_{n-1}})^* = \sum \det \frac{\partial(x_1, \dots, \widehat{x_i}, \dots, x_n)}{\partial(u_1, \dots, u_{n-1})} \vec{e}_{\wedge([n]-i)}^* \\ &= \sum (-1)^{n-i} \det \frac{\partial(x_1, \dots, \widehat{x_i}, \dots, x_n)}{\partial(u_1, \dots, u_{n-1})} \vec{e}_i. \end{aligned}$$

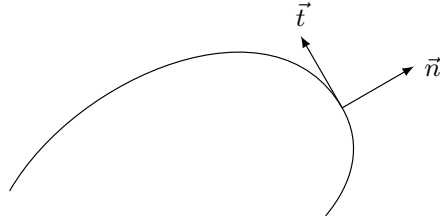
Then the *flux* of a vector field  $F = (f_1, \dots, f_n)$  along the oriented  $M$  is

$$\begin{aligned} \int_M F \cdot \vec{n} dV &= (-1)^{n-1} \int_A F \cdot \vec{a} du_1 \cdots du_k \\ &= \int_U \sum (-1)^{i-1} f_i \det \frac{\partial(x_1, \dots, \widehat{x_i}, \dots, x_n)}{\partial(u_1, \dots, u_{n-1})} du_1 \cdots du_k \\ &= \int_M \sum (-1)^{i-1} f_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n. \end{aligned} \quad (13.3.3)$$

If  $n = 3$ , then  $M$  is a surface in  $\mathbb{R}^3$ , and  $(-1)^{3-1}(\sigma_{u_1} \wedge \sigma_{u_2})^*$  is the usual cross product  $\sigma_{u_1} \times \sigma_{u_2}$ . Therefore (13.3.3) is the usual flux along an oriented surface in  $\mathbb{R}^3$ .

If  $n = 2$ , then  $M$  is a curve  $\phi(t)$  in  $\mathbb{R}^2$ . The property  $\det(\vec{n} \phi'(t)) > 0$  implies that the normal vector  $\vec{n}$  is the clockwise rotation of the unit length tangent vector  $\vec{t} = \frac{\phi'(t)}{\|\phi'(t)\|}$  by 90 degrees. The rotation is  $R(x, y) = (y, -x)$ , so that

$$\begin{aligned} \int_C (f, g) \cdot \vec{n} ds &= \int_C (f, g) \cdot R(\vec{t}) ds = \int_C R^{-1}(f, g) \cdot \vec{t} ds \\ &= \int_C (-g, f) \cdot d\vec{x} = \int_C f dy - g dx. \end{aligned}$$



**Figure 13.3.1.** Normal vector of oriented curve

**Example 13.3.3.** Consider the graph  $\sigma(\vec{u}) = (\vec{u}, f(\vec{u}))$ ,  $\vec{u} = (x_1, \dots, x_{n-1})$ , of a continuously differentiable function  $f$  on  $\mathbb{R}^{n-1}$ . By

$$\begin{aligned} &(\sigma_{x_1} \wedge \dots \wedge \sigma_{x_{n-1}})^* \\ &= ((\vec{e}_1 + f_{x_1} \vec{e}_n) \wedge \dots \wedge (\vec{e}_{n-1} + f_{x_{n-1}} \vec{e}_n))^* \\ &= (\vec{e}_1 \wedge \dots \wedge \vec{e}_{n-1})^* + \sum (-1)^{n-1-i} f_{x_i} (\vec{e}_1 \wedge \dots \wedge \widehat{\vec{e}_i} \wedge \dots \wedge \vec{e}_n)^* \\ &= \vec{e}_n - \sum f_{x_i} \vec{e}_i = (-f_{x_1}, \dots, -f_{x_{n-1}}, 1), \end{aligned}$$

the normal vector

$$\vec{n} = (-1)^{n-1} \frac{(-f_{x_1}, \dots, -f_{x_{n-1}}, 1)}{\sqrt{f_{x_1}^2 + \dots + f_{x_{n-1}}^2 + 1}}$$

points in the direction of  $x_n$  for odd  $n$  and opposite to the direction of  $x_n$  for even  $n$ .

Another way of finding  $\vec{n}$  is by using the three characteristic properties, similar to Example 13.2.6. First, for  $\vec{v} = (a_1, \dots, a_n)$ , we solve

$$\vec{v} \cdot \sigma_{x_i} = (v_1, \dots, v_n) \cdot (0, \dots, 1_{(i)}, \dots, 0, f_{x_i}) = v_i + v_n f_{x_i} = 0.$$

By taking  $v_n = 1$ , we get one solution  $\vec{v} = (-f_{x_1}, \dots, -f_{x_{n-1}}, 1)$ . Moreover,

$$\begin{aligned} \det(\vec{v} \sigma_{x_1} \dots \sigma_{x_{n-1}}) &= \det \begin{pmatrix} -f_{x_1} & 1 & 0 & \dots & 0 \\ -f_{x_2} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -f_{x_{n-1}} & 0 & 0 & \dots & 1 \\ 1 & f_{x_1} & f_{x_2} & \dots & f_{x_{n-1}} \end{pmatrix} \\ &= (-1)^{n-1} (f_{x_1}^2 + \dots + f_{x_{n-1}}^2 + 1). \end{aligned}$$



Therefore we get

$$\vec{n} = (-1)^{n-1} \frac{\vec{v}}{\|\vec{v}\|} = (-1)^{n-1} \frac{(-f_{x_1}, \dots, -f_{x_{n-1}}, 1)}{\sqrt{f_{x_1}^2 + \dots + f_{x_{n-1}}^2 + 1}}.$$

**Exercise 13.49.** Suppose  $U$  is an orthogonal transform of  $\mathbb{R}^n$ . Prove that if  $\vec{n}$  is the normal vector of a regular parameterization  $\sigma(\vec{u})$ , then  $(\det U)U(\vec{n})$  is the normal vector of the regular parameterization  $U(\sigma(\vec{u}))$ . Moreover, use this to show that the normal vector of the graph

$$\sigma_i(\vec{u}) = (x_1, \dots, x_{i-1}, f(\vec{u}), x_{i+1}, \dots, x_n), \quad \vec{u} = (x_1, \dots, \widehat{x_i}, \dots, x_n),$$

of a continuously differentiable function  $f$  is

$$(-1)^{i-1} \frac{(-f_{x_1}, \dots, -f_{x_{i-1}}, 1, -f_{x_{i+1}}, \dots, -f_{x_n})}{\sqrt{f_{x_1}^2 + \dots + \widehat{f_{x_i}^2} + \dots + f_{x_n}^2 + 1}}.$$

**Exercise 13.50.** Justify the normal vector in Exercise 13.49 by verifying the geometrical characterization.

**Exercise 13.51.** The unit sphere  $S^n$  in  $\mathbb{R}^{n+1}$  has the normal vector  $\vec{n} = \vec{x}$  at  $\vec{x} \in S^n$ . For each  $0 \leq i \leq n$ , find suitable rearrangements  $(x_{k_0}, \dots, \widehat{x_{k_i}}, \dots, x_{k_n})$  and  $(x_{l_0}, \dots, \widehat{x_{l_i}}, \dots, x_{l_n})$  of  $(x_0, \dots, \widehat{x_i}, \dots, x_n)$ , such that the parameterizations

$$\begin{aligned} \sigma_i^+(x_0, \dots, \widehat{x_i}, \dots, x_n) &= (x_{k_0}, \dots, x_{k_{i-1}}, \sqrt{1 - x_0^2 - \dots - \widehat{x_i^2} - \dots - x_n^2}, x_{k_{i+1}}, \dots, x_{k_n}), \\ \sigma_i^-(x_0, \dots, \widehat{x_i}, \dots, x_n) &= (x_{l_0}, \dots, x_{l_{i-1}}, -\sqrt{1 - x_0^2 - \dots - \widehat{x_i^2} - \dots - x_n^2}, x_{l_{i+1}}, \dots, x_{l_n}). \end{aligned}$$

cover  $S^n$  and induce the given normal vector. In particular, the parameterizations are compatibly oriented.

**Exercise 13.52.** Use (13.3.3) to show that the length of a curve  $C$  in  $\mathbb{R}^2$  given by an equation  $g(x, y) = c$  is  $\int_C \frac{-g_y dx + g_x dy}{\sqrt{g_x^2 + g_y^2}}$ , similar to the formula in Exercise 13.46. Extend the formula to the volume of an  $(n-1)$ -dimensional hypersurface in  $\mathbb{R}^n$  given by  $g(x_1, x_2, \dots, x_n) = c$ .

**Exercise 13.53.** Compute the integral.

1.  $\int_{S_R^{n-1}} (\vec{x} - \vec{a}) \cdot \vec{n} dV$ , the normal vector  $\vec{n}$  points outward of the sphere.
2.  $\int_{S_R^{n-1}} (a_1 x_1 + a_2 x_2 + \dots + a_n x_n) dx_2 \wedge dx_3 \wedge \dots \wedge dx_n$ , the orientation of the sphere is given by the outward normal vector.
3.  $\int_S dx_1 \wedge dx_2 \wedge dx_3 + dx_2 \wedge dx_3 \wedge dx_4 + \dots + dx_{n-2} \wedge dx_{n-1} \wedge dx_n$ ,  $S$  is the surface  $\sigma(u, v, w) = \rho(u) + \rho(v) + \rho(w)$ ,  $\rho(u) = (u, u^2, \dots, u^n)$ ,  $0 \leq u, v, w \leq a$ .

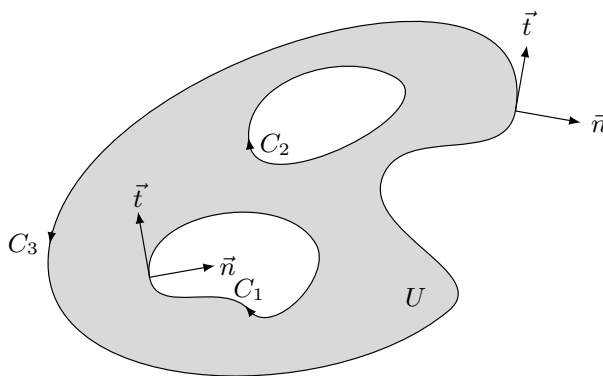
## 13.4 Green's Theorem

The Fundamental Theorem of Calculus relates the integration of a function on an interval to the values of the antiderivative at the end points of the integral. The Green's Theorem extends the fundamental theorem to the plane.

### 2-Dimensional Fundamental Theorem of Calculus in $\mathbb{R}^2$

A curve  $\phi: [a, b] \rightarrow \mathbb{R}^n$  is *simple* if it has no self intersection. It is *closed* if  $\phi(a) = \phi(b)$ . The concept of simple closed curve is independent of the choice of parameterization.

A simple closed curve in  $\mathbb{R}^2$  divides the plane into two connected pieces, one bounded and the other unbounded. Therefore we have *two sides* of a simple closed curve in  $\mathbb{R}^2$ . Suppose  $U \subset \mathbb{R}^2$  is a bounded subset with finitely many rectifiable simple closed curves  $C_1, C_2, \dots, C_k$  as the boundary. Then  $C_i$  has a *compatible orientation* such that  $U$  is “on the left” of  $C_i$ . This means that  $C_i$  has *counterclockwise* orientation if  $U$  is in the bounded side of  $C_i$ , and has *clockwise* orientation if  $U$  is in the unbounded side. Moreover, in case  $C_i$  is regular and differentiable, we have the *outward normal vector*  $\vec{n}$  along  $C_i$  pointing away from  $U$ . By rotating the outward normal vector by 90 degrees in the counterclockwise direction, we get the compatible direction  $\vec{t}$  of  $C_i$ .



**Figure 13.4.1.** Compatible orientation of boundary.

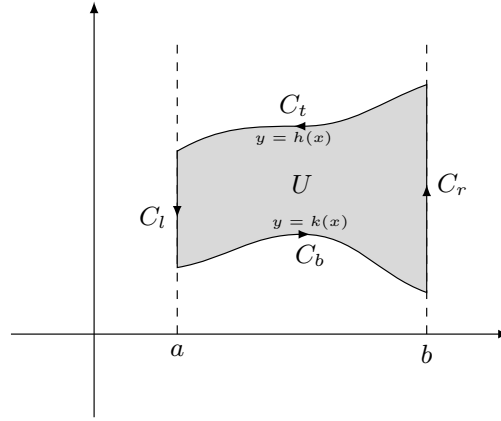
**Theorem 13.4.1** (Green's Theorem). Suppose  $U \subset \mathbb{R}^2$  is a region with compatibly oriented rectifiable boundary curve  $C$ . Then for any continuously differentiable functions  $f$  and  $g$ , we have

$$\int_C f dx + g dy = \int_U (-f_y + g_x) dx dy.$$

The curve  $C$  in the theorem is understood to be the union of simple closed curves  $C_1, C_2, \dots, C_k$ , each oriented in the compatible way. The integral along  $C$  is

$$\int_C = \int_{C_1} + \int_{C_2} + \dots + \int_{C_k}.$$

*Proof.* We only prove  $\int_C f dx = - \int_U f_y dx dy$ . By exchanging  $x$  and  $y$  (which reverses the orientation), this also gives  $\int_C g dy = \int_U g_x dx dy$ . Adding the two equalities gives the whole formula.



**Figure 13.4.2.** *Green's Theorem for a special case.*

We first consider the special case that

$$U = \{(x, y) : x \in [a, b], h(x) \geq y \geq k(x)\}$$

is a region between the graphs of functions  $h(x)$  and  $k(x)$  with bounded variations. We have

$$\int_U f_y dx dy = \int_a^b \left( \int_{k(x)}^{h(x)} f_y(x, y) dy \right) dx \quad (\text{Fubini Theorem})$$

$$= \int_a^b (f(x, h(x)) - f(x, k(x))) dx. \quad (\text{Fundamental Theorem})$$

On the other hand, the boundary  $C$  consists of four segments with the following orientation compatible parameterizations.

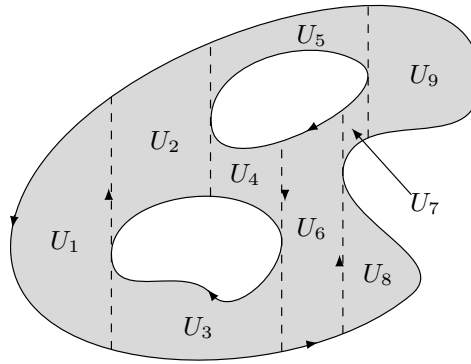
$$\begin{aligned} C_b: \phi(t) &= (t, k(t)), & t &\in [a, b] \\ C_r: \phi(t) &= (b, t), & t &\in [k(b), h(b)] \\ C_t: \phi(t) &= (-t, h(-t)), & t &\in [-b, -a] \\ C_l: \phi(t) &= (a, -t), & t &\in [-h(a), -k(a)] \end{aligned}$$

Then

$$\begin{aligned}
 \int_C f dx &= \left( \int_{C_b} + \int_{C_r} + \int_{C_t} + \int_{C_l} \right) f dx \\
 &= \int_a^b f(t, k(t)) dt + \int_{k(b)}^{h(b)} f(b, t) db \\
 &\quad + \int_{-b}^{-a} f(-t, h(-t)) d(-t) + \int_{-h(a)}^{-k(a)} f(a, -t) da \\
 &= \int_a^b f(x, k(x)) dx + 0 + \int_b^a f(x, h(x)) dx + 0.
 \end{aligned}$$

Therefore we have  $\int_U f_y dx dy = - \int_C f dx$ .

In general, the region  $U$  can often be divided by vertical lines into several special regions  $U_j$  studied above. Let  $C_j$  be the compatibly oriented boundary of  $U_j$ . Then the sum of  $\int_{C_j} f dx$  is  $\int_C f dx$ , because the integrations along the vertical lines are all zero. Moreover, the sum of  $\int_{U_j} f_y dx dy$  is  $\int_U f_y dx dy$ . This proves the equality  $\int_C f dx = - \int_U f_y dx dy$  for the more general regions.



**Figure 13.4.3.** Divide into special cases: Orientations for  $U_1, U_6$  given.

Now consider the most general case that the boundary curves are assumed to be only rectifiable. To simplify the presentation, we will assume that  $U$  has only one boundary curve  $C$ . Let  $P$  be a partition of  $C$ . Let  $L$  be the union of straight line segments  $L_i$  connecting the partition points  $\vec{x}_{i-1}$  and  $\vec{x}_i$ . Let  $U'$  be the region enclosed by the curve  $L$ . Then  $U'$  can be divided by vertical lines into special

regions. As explained above, we have

$$\int_L f dx = - \int_{U'} f_y dx dy.$$

We will show that, as  $\|P\| = \max \mu(C_i) \rightarrow 0$ , the left side converges to  $\int_C f dx$ , and the right side converges to  $-\int_U f_y dx dy$ . Then the general case is proved.

For any  $\epsilon > 0$ , by the uniform continuity of  $f$  on compact  $C \cup L$ , there is  $\delta > 0$ , such that  $\vec{x}, \vec{y} \in C \cup L$  and  $\|\vec{x} - \vec{y}\| < \delta$  imply  $|f(\vec{x}) - f(\vec{y})| < \epsilon$ . If  $\|P\| < \delta$ , then for  $\vec{x} \in C_i \cup L_i$ , we have  $|f(\vec{x}) - f(\vec{x}_i)| < \epsilon$ . This implies

$$\left| \left( \int_{C_i} - \int_{L_i} \right) f(\vec{x}) dx \right| = \left| \left( \int_{C_i} - \int_{L_i} \right) (f(\vec{x}) - f(\vec{x}_i)) dx \right| \leq \epsilon \mu(C_i).$$

Therefore

$$\left| \int_C f dx - \int_L f dx \right| \leq \sum \left| \left( \int_{C_i} - \int_{L_i} \right) f dx \right| \leq \sum \epsilon \mu(C_i) = \epsilon \mu(C).$$

This proves  $\lim_{\|P\| \rightarrow 0} \int_L f dx = \int_C f dx$ .

On the other hand, let  $M$  be the upper bound of the continuous function  $\|\nabla f\|_2 = \sqrt{f_x^2 + f_y^2}$  on the compact region enclosed by  $C$  and  $L$ . Let  $B_j$  be the region bounded by  $C_j$  and  $L_j$ . Then  $B_j$  is contained in the ball of center  $\vec{x}_i$  and radius  $\mu(C_i)$ . Therefore the area  $\mu(B_j) \leq \pi \mu(C_i)^2 \leq \pi \delta \mu(C_i)$ , and we have

$$\begin{aligned} \left| \int_U f dx dy - \int_{U'} f dx dy \right| &\leq \sum \int_{B_j} |f| dx dy \\ &\leq \sum M \mu(B_j) \leq 2\pi \delta M \sum \mu(C_i) = 2\pi \delta M \mu(C). \end{aligned}$$

This proves  $\lim_{\|P\| \rightarrow 0} \int_{U'} f dx dy = \int_U f dx dy$ .  $\square$

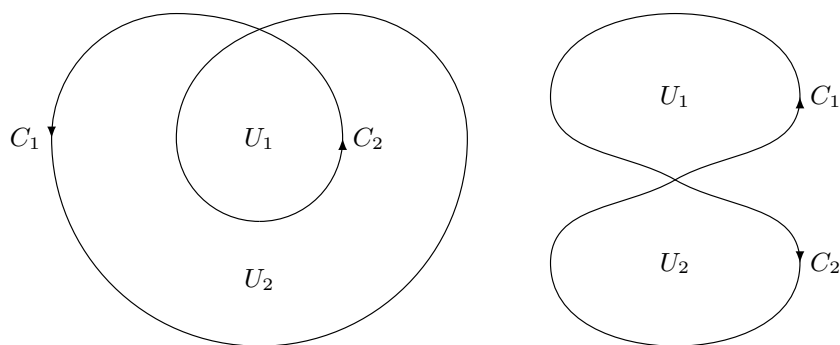
A closed but not necessarily simple curve may enclose some parts of the region more than once, and the orientation of the curve may be the opposite of what is supposed to be. For example, in Figure 13.4.4, the left curve intersects itself once, which divides the curve into the outside part  $C_1$  and the inside part  $C_2$ . Then

$$\begin{aligned} \int_C f dx + g dy &= \left( \int_{C_1} + \int_{C_2} \right) f dx + g dy = \left( \int_{U_1 \cup U_2} + \int_{U_1} \right) (-f_y + g_x) dx dy \\ &= \left( 2 \int_{U_1} + \int_{U_2} \right) (-f_y + g_x) dx dy. \end{aligned}$$

For the curve on the right, we have

$$\int_C f dx + g dy = \left( \int_{U_1} - \int_{U_2} \right) (-f_y + g_x) dx dy$$

In general, Green's Theorem still holds, as long as the regions are counted in the corresponding way. In fact, this remark is already used in the proof of Green's Theorem above, since the curve  $L$  in the proof may not be simple.



**Figure 13.4.4.** Closed but not simple curves.

**Example 13.4.1.** The area of the region  $U$  enclosed by a simple closed curve  $C$  is

$$\int_U dx dy = \int_C x dy = - \int_C y dx.$$

For example, suppose  $C$  is the counterclockwise unit circle arc from  $(1, 0)$  to  $(0, 1)$ . To compute  $\int_C x dy$ , we add  $C_1 = [0, 1] \times 0$  in the rightward direction and  $C_2 = 0 \times [0, 1]$  in the downward direction. Then the integral  $\left( \int_C + \int_{C_1} + \int_{C_2} \right) x dy$  is the area  $\frac{\pi}{4}$  of the quarter unit disk. Since  $\int_{C_1} x dy = \int_0^1 x d0 = 0$  and  $\int_{C_2} x dy = \int_1^0 0 dy = 0$ , we conclude that  $\int_C x dy = \frac{\pi}{4}$ .

**Exercise 13.54.** Compute the area.

1. Region enclosed by the astroid  $x^{\frac{3}{2}} + y^{\frac{3}{2}} = 1$ .
2. Region enclosed by the cardioid  $r = 2a(1 + \cos \theta)$ .
3. Region enclosed by the parabola  $(x + y)^2 = x$  and the  $x$ -axis.

**Exercise 13.55.** Suppose  $\phi(t) = (x(t), y(t)) : [a, b] \rightarrow \mathbb{R}^2$  is a curve, such that the rotation from  $\phi$  to  $\phi'$  is counterclockwise. Prove that the area swept by the line connecting the origin and  $\phi(t)$  as  $t$  moves from  $a$  to  $b$  is  $\frac{1}{2} \int_a^b (xy' - yx') dt$ .

**Exercise 13.56.** Kepler's Law of planet motion says that the line from the sun to the planet sweeps out equal areas in equal intervals of time. Derive Kepler's law from Exercise 13.55 and Newton's second law of motion:  $\phi'' = c \frac{\phi}{\|\phi\|_2^3}$ , where  $c$  is a constant determined by the mass of the sun.

Exercise 13.57. Extend integration by parts by finding a relation between  $\int_U (uf_x + vg_y) dx dy$  and  $\int_U (u_x f + v_y g) dx dy$ .

Exercise 13.58. The *divergence* of a vector field  $F = (f, g)$  on  $\mathbb{R}^2$  is

$$\operatorname{div} F = f_x + g_y.$$

Prove that

$$\int_C F \cdot \vec{n} ds = \int_U \operatorname{div} F dA,$$

where  $\vec{n}$  is the outward normal vector pointing away from  $U$ .

Exercise 13.59. The Laplacian of a two variable function  $f$  is  $\Delta f = f_{xx} + f_{yy} = \operatorname{div} \nabla f$ . Prove Green's identities

$$\int_U f \Delta g dA = \int_C f \nabla g \cdot \vec{n} ds - \int_U \nabla f \cdot \nabla g dA,$$

and

$$\int_U (f \Delta g - g \Delta f) dA = \int_C (f \nabla g - g \nabla f) \cdot \vec{n} ds.$$

### Potential: Antiderivative on $\mathbb{R}^2$

If  $f_y = g_x$  on a region  $U \subset \mathbb{R}^2$  and  $C$  is the boundary curve, then Green's Theorem tells us  $\int_C f dx + g dy = 0$ . The conclusion can be interpreted and utilized in different ways.

**Theorem 13.4.2.** Suppose  $f$  and  $g$  are continuous functions on an open subset  $U \subset \mathbb{R}^2$ . Then the following are equivalent.

1. The integral  $\int_C f dx + g dy$  along an oriented rectifiable curve  $C$  in  $U$  depends only on the beginning and end points of  $C$ .
2. There is a differentiable function  $\varphi$  on  $U$ , such that  $\varphi_x = f$  and  $\varphi_y = g$ .

Moreover, if  $U$  is simply connected and  $f$  and  $g$  are continuously differentiable, then the above is also equivalent to

3.  $f_y = g_x$  on  $U$ .

The reason for the second statement to imply the first already appeared in Examples 13.1.6. Suppose  $f = \varphi_x$  and  $g = \varphi_y$  for differentiable  $\varphi$ , and  $C$  is parameterized by differentiable  $\phi(t) = (x(t), y(t))$ ,  $t \in [a, b]$ . Then

$$\int_C f dx + g dy = \int_a^b (\varphi_x x' + \varphi_y y') dt = \int_a^b \frac{d\varphi(\phi(t))}{dt} dt = \varphi(\phi(b)) - \varphi(\phi(a)).$$

The right side depends only on the beginning point  $\phi(a)$  and the end point  $\phi(b)$  of the curve. For the proof in case  $C$  is only rectifiable, see Exercise 13.65.

To prove the converse that the first statement implies the second, we note that  $f = \varphi_x$  and  $g = \varphi_y$  means that the vector field  $(f, g)$  is the gradient  $\nabla\varphi$ . It also means that the 1-form  $f dx + g dy$  is the differential  $d\varphi$ . The function  $\varphi$ , called a *potential* of the vector field  $(f, g)$  or the 1-form  $f dx + g dy$ , is actually the two-variable antiderivative of the vector field or the 1-form.

On  $\mathbb{R}^1$ , any continuous single variable function  $f(x)$  has antiderivative

$$\varphi(x) = \int_a^x f(t) dt.$$

On  $\mathbb{R}^2$ , we may imitate the single variable case and define

$$\varphi(x, y) = \int_{(a,b)}^{(x,y)} f dx + g dy. \quad (13.4.1)$$

The problem is that the integral should be along a path connecting  $(a, b)$  to  $(x, y)$ , but there are many choices of such paths. In contrast, the integral for the single variable case is simply over the interval  $[a, x]$ , the “unique path” connecting the two points.

Now we understand that the first statement of Theorem 13.4.2 means exactly that the function (13.4.1) is well defined. It remains to show that the function satisfies  $\varphi_x = f$  and  $\varphi_y = g$ . Specifically, we will prove that

$$L(x, y) = \varphi(x_0, y_0) + f(x_0, y_0)(x - x_0) + g(x_0, y_0)(y - y_0)$$

is the linear approximation of  $\varphi$  near  $(x_0, y_0) \in U$ . By the continuity of  $f$  and  $g$  and the openness of  $U$ , for any  $\epsilon > 0$ , we can find  $\delta > 0$ , such that  $\|(x, y) - (x_0, y_0)\|_2 < \delta$  implies  $(x, y) \in U$  and  $\|(f(x, y), g(x, y)) - (f(x_0, y_0), g(x_0, y_0))\|_2 < \epsilon$ . For such  $(x, y)$ , let  $C$  be a path in  $U$  connecting  $(a, b)$  to  $(x_0, y_0)$  and let  $I$  be the straight line  $I$  connecting  $(x_0, y_0)$  to  $(x, y)$ . Then

$$\varphi(x, y) - \varphi(x_0, y_0) = \left( \int_{C \cup I} - \int_C \right) f dx + g dy = \int_I (f dx + g dy).$$

Then we have

$$\begin{aligned} |\varphi(x, y) - L(x, y)| &= \left| \int_I (f(x, y) - f(x_0, y_0)) dx + (g(x, y) - g(x_0, y_0)) dy \right| \\ &\leq \sup_{(x,y) \in I} \|(f(x, y) - f(x_0, y_0), g(x, y) - g(x_0, y_0))\|_2 \mu(I) \\ &\leq \epsilon \|(x, y) - (x_0, y_0)\|_2. \end{aligned}$$

Here the first inequality follows from Exercise 13.32. This proves the linear approximation.

Now we come to the third statement of Theorem 13.4.2. The third statement is a local property of  $f$  and  $g$ , because the partial derivatives depend only on the



values of the functions near individual points. In contrast, the first two statements are global properties of  $f$  and  $g$ . We also note that the equivalence of the first two statements do not rely on Green's Theorem.

The global property often implies the local property, without any extra condition. In our case, given the first statement, for the boundary  $C$  of any small disk  $B \subset U$ , by Green's Theorem, we have

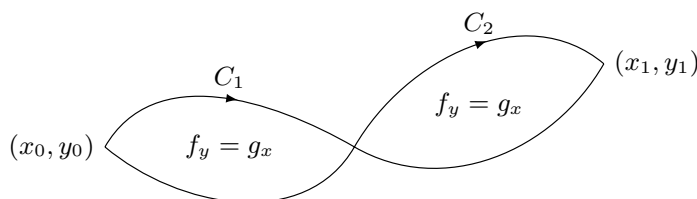
$$\int_B (-f_y + g_x) dx dy = \int_C f dx + g dy = 0.$$

Here the first statement implies that the integral along a curve  $C$  with the same beginning and end points must be 0. Since the integral on the left is always 0 for all  $B$ , we conclude that the integrand  $-f_y + g_x = 0$  everywhere. This is the third statement.

Now we prove the converse that the third statement implies the first. Suppose  $C_1$  and  $C_2$  are two curves connecting  $(x_0, y_0)$  to  $(x_1, y_1)$  in  $\mathbb{R}^2$ . Then  $C_1 \cup (-C_2)$  is an oriented closed curve. If the region  $B$  enclosed by  $C_1 \cup (-C_2)$  is contained in  $U$ , then we have  $f_y = g_x$  on  $B$ , so that

$$\left( \int_{C_1} - \int_{C_2} \right) f dx + g dy = \int_{C_1 \cup (-C_2)} f dx + g dy = \int_B (-f_y + g_x) dx dy = 0.$$

Here the second equality is Green's Theorem.



**Figure 13.4.5.** Integrals along  $C_1$  and  $C_2$  are the same.

For the argument to always work, however, we need the region  $B$  enclosed by any closed curve  $C_1 \cup (-C_2) \subset U$  to be contained in  $U$ . This means that  $U$  has no holes and is the *simply connected* condition on  $U$  in Theorem 13.4.2. The rigorous definition for a subset  $U \subset \mathbb{R}^n$  to be simply connected is that any continuous map  $S^1 \rightarrow U$  from the unit circle extends to a continuous map  $B^2 \rightarrow U$  from the unit disk.

**Example 13.4.2.** The integral of  $ydx + xdy$  in Example 13.1.6 is independent of the choice of the paths because the equality  $\frac{\partial y}{\partial y} = \frac{\partial x}{\partial x}$  holds on the whole plane, which is simply connected. The potential of the 1-form is  $xy$ , up to adding constants.

In Example 13.1.7, the integrals of  $xydx + (x + y)dy$  along the three curves are different. Indeed, we have  $(xy)_y = x \neq (x + y)_x = 1$ , and the 1-form has no potential.

**Example 13.4.3.** The vector field  $\frac{1}{y^2}(xy^2 + y, 2y - x)$  is defined on  $y > 0$  and  $y < 0$ , both simply connected. It satisfies  $\frac{d}{dy} \left( \frac{xy^2 + y}{y^2} \right) = -\frac{1}{y^2} = \frac{d}{dx} \left( \frac{2y - x}{y^2} \right)$ . Therefore it has a potential function  $\varphi$ . By  $\varphi_x = \frac{y(xy + 1)}{y^2}$ , we get  $\varphi = \int \frac{y(xy + 1)}{y^2} dx + \vartheta(y) = \frac{x^2}{2} + \frac{x}{y} + \vartheta(y)$ . Then by  $\varphi_y = -\frac{x}{y^2} + \vartheta'(y) = \frac{2y - x}{y^2}$ , we get  $\vartheta'(y) = \frac{2}{y}$ , so that  $\vartheta(y) = 2 \log |y| + c$ . The potential function is

$$\varphi = \frac{x^2}{2} + \frac{x}{y} + 2 \log |y| + c.$$

In particular, we have

$$\int_{(1,1)}^{(2,2)} \frac{(xy^2 + y)dx + (2y - x)dy}{y^2} = \left( \frac{x^2}{2} + \frac{x}{y} + 2 \log |y| \right)_{(1,1)}^{(2,2)} = \frac{3}{2} + \log 2.$$

The example shows that the formula (13.4.1) may not be the most practical way of computing the potential.

**Example 13.4.4.** The 1-form  $\frac{ydx - xdy}{x^2 + y^2}$  satisfies

$$f_y = \left( \frac{y}{x^2 + y^2} \right)_y = \frac{x^2 - y^2}{(x^2 + y^2)^2} = g_x = \left( \frac{-x}{x^2 + y^2} \right)_x$$

on  $\mathbb{R}^2 - (0, 0)$ , which is unfortunately not simply connected. Let  $U$  be obtained by removing the non-positive  $x$ -axis  $(-\infty, 0] \times 0$  from  $\mathbb{R}^2$ . Then  $U$  is simply connected, and Theorem 13.4.2 can be applied. The potential for  $\frac{ydx - xdy}{x^2 + y^2}$  is  $\varphi = -\theta$ , where  $-\pi < \theta < \pi$  is the angle in the polar coordinate. The potential can be used to compute

$$\int_{(1,0)}^{(0,1)} \frac{ydx - xdy}{x^2 + y^2} = -\theta(0, 1) + \theta(1, 0) = -\frac{\pi}{2} + 0 = -\frac{\pi}{2}$$

for any curve connecting  $(1, 0)$  to  $(0, 1)$  that does not intersect the non-positive  $x$ -axis. The unit circle arc  $C_1: \phi_1(t) = (\cos t, \sin t)$ ,  $0 \leq t \leq \frac{\pi}{2}$ , and the straight line  $C_2: \phi_2(t) = (1 - t, t)$ ,  $0 \leq t \leq 1$ , are such curves.

On the other hand, consider the unit circle arc  $C_3: \phi_3(t) = (\cos t, -\sin t)$ ,  $0 \leq t \leq \frac{3\pi}{2}$ , connecting  $(1, 0)$  to  $(0, 1)$  in the clockwise direction. To compute the integral along the curve, let  $V$  be obtained by removing the non-negative diagonal  $\{(x, x): x \geq 0\}$  from  $\mathbb{R}^2$ . The 1-form still has the potential  $\varphi = -\theta$  on  $V$ . The crucial difference is that the range for  $\theta$  is changed to  $\frac{\pi}{4} < \theta < \frac{9\pi}{4}$ . Therefore

$$\int_{(1,0)}^{(0,1)} \frac{ydx - xdy}{x^2 + y^2} = -\theta(0, 1) + \theta(1, 0) = -\frac{\pi}{2} + 2\pi = \frac{3\pi}{2}$$

for any curve connecting  $(1, 0)$  to  $(0, 1)$  that does not intersect the non-negative diagonal.

In general, the integral  $\int_{(1,0)}^{(0,1)} \frac{ydx - xdy}{x^2 + y^2}$  depends only on how the curve connecting  $(1, 0)$  to  $(0, 1)$  goes around the origin, the only place where  $f_y = g_x$  fails.

Exercise 13.60. Explain the computations in Exercise 13.25 by Green's Theorem.

Exercise 13.61. Use potential function to compute the integral.

1.  $\int_C 2x dx + y dy$ ,  $C$  is the straight line connecting  $(2, 0)$  to  $(0, 2)$ .
2.  $\int_C \frac{x dx + y dy}{x^2 + y^2}$ ,  $C$  is the circular arc connecting  $(2, 0)$  to  $(0, 2)$ .
3.  $\int_C \frac{y dx - x dy}{ax^2 + by^2}$ ,  $C$  is the unit circle in counterclockwise direction.
4.  $\int_C \frac{(x-y)dx + (x+y)dy}{x^2 + y^2}$ ,  $C$  is the elliptic arc  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$  in the upper half plane connecting  $(a, 0)$  to  $(-a, 0)$ .
5.  $\int_C (e^x \sin 2y - y)dx + (2e^x \cos 2y - 1)dy$ ,  $C$  is the circular arc connecting  $(0, 1)$  to  $(1, 0)$  in clockwise direction.
6.  $\int_C (2xy^3 - 3y^2 \cos x)dx + (-6y \sin x + 3x^2 y^2)dy$ ,  $C$  is the curve  $2x = \pi y^2$  connecting  $(0, 0)$  to  $(\frac{\pi}{2}, 1)$ .

Exercise 13.62. Find  $p$  so that the vector fields  $\frac{(x, y)}{(x^2 + y^2)^p}$  and  $\frac{(-y, x)}{(x^2 + y^2)^p}$  have potentials. Then find the potentials.

Exercise 13.63. Compute the integral  $\int_C g(xy)(y dx + x dy)$ , where  $C$  is the straight line connecting  $(2, 3)$  to  $(1, 6)$ .

Exercise 13.64. Explain that the potential function is unique up to adding constants.

Exercise 13.65. We only proved that the second statement of Theorem 13.4.2 implies the first for the case that the curve  $C$  has differentiable parameterization. Extend the proof to the case  $C$  is divided into finitely many segments, such that each segment has differentiable parameterization. Then further extend to general rectifiable  $C$  by using the idea of the proof of Theorem 13.4.1.

## Integral Along Closed Curve

If  $f_y = g_x$  is satisfied, then the integral along a curve is “largely” dependent only on the end points. In fact, the value is only affected by the existence of holes, as illustrated by Example 13.4.4.

The open subset in Figure 13.4.6 has two holes, which are enclosed by simple closed curves  $C_1$  and  $C_2$  in  $U$  oriented in counterclockwise direction (*opposite* to the direction adopted in Green's Theorem). The closed curve  $C$  in  $U$  may be divided into four parts, denoted  $C_{[1]}$ ,  $C_{[2]}$ ,  $C_{[3]}$ ,  $C_{[4]}$ . The unions of oriented closed curves  $C_{[1]} \cup C_{[3]} \cup (-C_1)$ ,  $C_{[2]} \cup (-C_1)$ ,  $(-C_{[4]}) \cup (-C_2)$  also enclose subsets  $U_1, U_2, U_3 \subset U$ .

If  $f_y = g_y$  on  $U$ , then by Green's Theorem, we have

$$\begin{aligned} \left( \int_{C_{[1]}} + \int_{C_{[3]}} - \int_{C_1} \right) f dx + g dy &= \int_{C_{[1] \cup C_{[3]} \cup (-C_1)} f dx + g dy \\ &= \int_{U_1} (-f_y + g_x) dx dy = 0. \end{aligned}$$

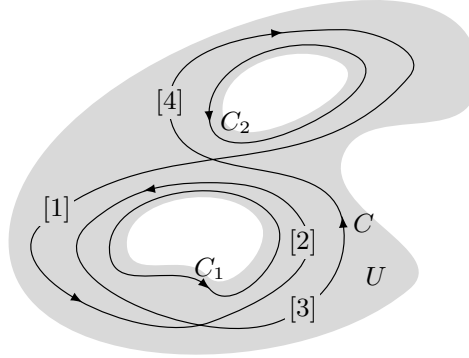
Similar argument can be made for the other two unions, and we get

$$\int_{C_{[1]}} + \int_{C_{[3]}} = \int_{C_1}, \quad \int_{C_{[2]}} = \int_{C_1}, \quad \int_{C_{[4]}} = -\int_{C_2}.$$

Adding the three equalities together, we get

$$\int_C f dx + g dy = \left( 2 \int_{C_1} - \int_{C_2} \right) f dx + g dy.$$

We remark that the coefficient 2 for  $C_1$  means that  $C$  wraps around  $C_1$  twice in the same direction, and the coefficient  $-1$  for  $C_2$  means that  $C$  wraps around  $C_2$  once in the opposite direction.



**Figure 13.4.6.** Closed curve in open subset with holes.

In general, suppose  $U$  has finitely many holes. We enclose these holes with closed curves  $C_1, \dots, C_k$ , all in counterclockwise orientation. Then any closed curve  $C$  in  $U$  wraps around the  $i$ -th hole  $n_i$  times. The sign of  $n_i$  is positive when the wrapping is counterclockwise (same as  $C_i$ ) and is negative when the wrapping is clockwise (opposite to  $C_i$ ). We say  $C$  is *homologous* to  $n_1 C_1 + \dots + n_k C_k$ , and in case  $f_y = g_x$  on  $U$ , we have

$$\int_C f dx + g dy = \left( n_1 \int_{C_1} + \dots + n_k \int_{C_k} \right) f dx + g dy.$$

**Example 13.4.5.** In Example 13.4.4, the 1-form  $\frac{ydx - xdy}{x^2 + y^2}$  satisfies  $f_y = g_x$  on  $U = \mathbb{R}^2 - (0, 0)$ . Since the unit circle  $C$  in the counterclockwise direction encloses the only hole of  $U$ , we have

$$\int_C \frac{ydx - xdy}{x^2 + y^2} = \int_0^{2\pi} \frac{\sin t(-\sin t)dt - \cos t \cos t dt}{\cos^2 t + \sin^2 t} = -2\pi.$$

If  $C_1$  is a curve on the first quadrangle connecting  $(1, 0)$  to  $(0, 1)$  and  $C_2$  is a curve on the second, third and fourth quadrangles connecting the two points, then  $C_1 \cup (-C_2)$  is homologous to  $C$ , and we have

$$\left( \int_{C_1} - \int_{C_2} \right) \frac{ydx - xdy}{x^2 + y^2} = \int_C \frac{ydx - xdy}{x^2 + y^2} = -2\pi.$$

Therefore

$$\int_{C_1} \frac{ydx - xdy}{x^2 + y^2} = \int_{C_2} \frac{ydx - xdy}{x^2 + y^2} - 2\pi.$$

**Exercise 13.66.** Study how the integral of the 1-form  $\frac{ydx - xdy}{x^2 + xy + y^2}$  depends on the curves.

**Exercise 13.67.** Study how the integral of the 1-form

$$\omega = \frac{(x^2 - y^2 - 1)dx + 2xydy}{((x - 1)^2 + y^2)((x + 1)^2 + y^2)}$$

depends on the curves. Note that if  $C_\epsilon$  is the counterclockwise circle of radius  $\epsilon$  around  $(1, 0)$ , then  $\int_{C_{\epsilon_0}} \omega = \lim_{\epsilon \rightarrow 0} \int_{C_\epsilon} \omega$ .

## 13.5 Stokes' Theorem

Green's Theorem is the Fundamental Theorem of Calculus for 2-dimensional body in  $\mathbb{R}^2$ . The theorem can be extended to 2-dimensional surface in  $\mathbb{R}^n$ .

### 2-Dimensional Fundamental Theorem of Calculus in $\mathbb{R}^3$

Let  $S \subset \mathbb{R}^3$  be an oriented surface. Let

$$\sigma(u, v) = (x(u, v), y(u, v), z(u, v)): U \subset \mathbb{R}^2 \rightarrow S \subset \mathbb{R}^3$$

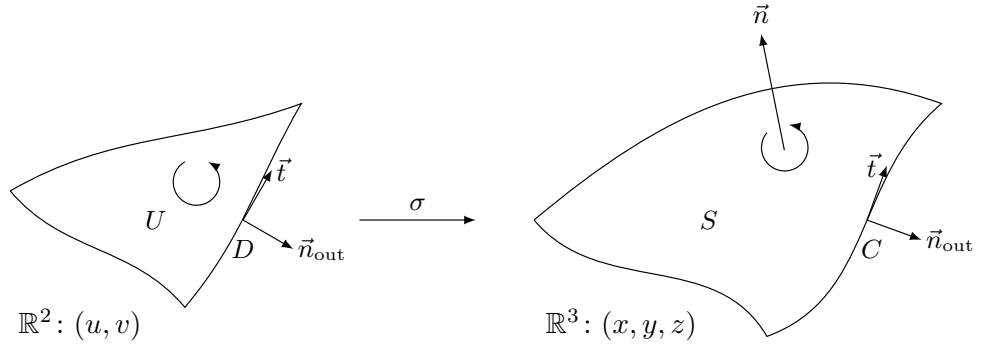
be an orientation compatible parameterization of  $S$ . Then the boundary  $C$  of  $S$  corresponds to the boundary  $D$  of  $U$ . Recall that  $D$  should be oriented in such a way that  $U$  is "on the left" of  $D$ . Correspondingly,  $C$  should also be oriented in such a way that  $S$  is "on the left" of  $C$ .

The orientation of  $C$  can be characterized by the outward normal vector  $\vec{n}_{\text{out}}$  along  $C$ , by requiring that the rotation from  $\vec{n}_{\text{out}}$  to the tangent vector  $\vec{t}$  of  $C$  is comparable to the rotation from  $\sigma_u$  to  $\sigma_v$ . Note that both  $\{\vec{n}_{\text{out}}, \vec{t}\}$  and  $\{\sigma_u, \sigma_v\}$  are

bases of the tangent space of  $S$  at boundary point. The same rotation requirement means that the matrix between the two bases has positive determinant, or

$$\sigma_u \times \sigma_v = \lambda \vec{n}_{\text{out}} \times \vec{t}, \quad \lambda > 0.$$

Here  $\lambda$  is the determinant of the matrix for expressing  $\{\sigma_u, \sigma_v\}$  in terms of  $\{\vec{n}_{\text{out}}, \vec{t}\}$ . This is also equivalent to  $\vec{n}_{\text{out}} \times \vec{t}$  being the normal vector  $\vec{n}$  of the surface  $S$ , and equivalent to  $\det(\vec{n}_{\text{out}} \sigma_u \sigma_v) > 0$ .



**Figure 13.5.1.** Boundary orientation compatible with the outward normal vector.

If  $F = (f, g, h)$  is a continuously differentiable vector field on  $\mathbb{R}^3$ , then

$$\begin{aligned} \int_C f dx + g dy + h dz &= \int_D (fx_u + gy_u + hz_u) du + (fx_v + gy_v + hz_v) dv \\ &= \int_U ((fx_v + gy_v + hz_v)_u - (fx_u + gy_u + hz_u)_v) dudv \\ &= \int_U (f_u x_v - f_v x_u + g_u y_v - g_v y_u + h_u z_v - h_v z_u) dudv. \end{aligned}$$

The first equality is due to

$$\begin{aligned} \int_C f dx &= \int_a^b f(x(t), y(t), z(t)) x'(t) dt \\ &= \int_a^b f(x, y, z) (x_u u'(t) + y_u u'(t)) dt \\ &= \int_D f x_u du + f x_v dv, \end{aligned}$$

and the similar equalities for  $\int_C g dy$  and  $\int_C h dz$ . The second equality is Green's Theorem.

By

$$\begin{aligned} f_u x_v - f_v x_u &= (f_x x_u + f_y y_u + f_z z_u) x_v - (f_x x_v + f_y y_v + f_z z_v) x_u \\ &= -f_y \det \frac{\partial(x, y)}{\partial(u, v)} + f_z \det \frac{\partial(z, x)}{\partial(u, v)}, \end{aligned}$$

we have

$$\begin{aligned}\int_U (f_u x_v - f_v x_u) du dv &= \int_U -f_y \det \frac{\partial(x, y)}{\partial(u, v)} du dv + f_z \det \frac{\partial(z, x)}{\partial(u, v)} du dv \\ &= \int_S -f_y dx \wedge dy + f_z dz \wedge dx.\end{aligned}$$

Combined with the similar equalities for  $g$  and  $h$ , we get the following result.

**Theorem 13.5.1** (Stokes' Theorem). *Suppose  $S \subset \mathbb{R}^3$  is an oriented surface with compatibly oriented boundary curve  $C$ . Then for any continuously differentiable  $f, g, h$  along  $S$ , we have*

$$\int_C f dx + g dy + h dz = \int_S (g_x - f_y) dx \wedge dy + (h_y - g_z) dy \wedge dz + (f_z - h_x) dz \wedge dx.$$

Using the symbol

$$\nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right),$$

the gradient of a function  $f$  can be formally denoted as

$$\text{grad} f = \nabla f = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right).$$

Define the *curl* of a vector field  $F = (f, g, h)$  to be

$$\text{curl} F = \nabla \times F = \left( \frac{\partial h}{\partial y} - \frac{\partial g}{\partial z}, \frac{\partial f}{\partial z} - \frac{\partial h}{\partial x}, \frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right).$$

Then Stokes' Theorem can be written as

$$\int_C F \cdot d\vec{x} = \int_S \text{curl} F \cdot \vec{n} dA.$$

**Example 13.5.1.** Suppose  $C$  is the circle given by  $x^2 + y^2 + z^2 = 1$  and  $x + y + z = r$ , with the counterclockwise orientation when viewed from the direction of the  $x$ -axis. We would like to compute the integral  $\int_C (ax + by + cz + d) dx$ . Let  $S$  be the disk given by  $x^2 + y^2 + z^2 \leq 1$  and  $x + y + z = r$ , with the normal direction given by  $(1, 1, 1)$ . Then by Stokes' Theorem and the fact that the radius of  $S$  is  $\sqrt{1 - \frac{r^2}{3}}$ , we have

$$\int_C (ax + by + cz + d) dx = \int_S -b dx \wedge dy + c dz \wedge dx = \int_S \frac{1}{\sqrt{3}} (-b + c) dA = \frac{c - b}{\sqrt{3}} \pi \left( 1 - \frac{r^2}{3} \right).$$

**Example 13.5.2.** Faraday observed that a changing magnetic field  $B$  induces an electric field  $E$ . More precisely, *Faraday's induction law* says that the rate of change of the flux of the magnetic field through a surface  $S$  is the negative of the integral of the electric field along the boundary  $C$  of the surface  $S$ . The law can be summarised as the equality

$$-\int_C E \cdot d\vec{x} = \frac{d}{dt} \int_S B \cdot \vec{n} dA.$$

By Stokes' Theorem, the left side is  $-\int_S \operatorname{curl} E \cdot \vec{n} dA$ . Since the equality holds for any surface  $S$ , we conclude the differential version of Faraday's law

$$-\operatorname{curl} E = \frac{\partial B}{\partial t}.$$

This is one of Maxwell's equations for electromagnetic fields.

**Exercise 13.68.** Compute  $\int_C y^2 dx + (x+y)dy + yzdz$ , where  $C$  is the ellipse  $x^2 + y^2 = 2$ ,  $x + y + z = 2$ , with clockwise orientation as viewed from the origin.

**Exercise 13.69.** Suppose  $C$  is any closed curve on the sphere  $x^2 + y^2 + z^2 = R^2$ . Prove that  $\int_C (y^2 + z^2)dx + (z^2 + x^2)dy + (x^2 + y^2)dz = 0$ . In general, what is the condition for  $f$ ,  $g$ ,  $h$  so that  $\int_C f dx + g dy + h dz = 0$  for any closed curve  $C$  on any sphere centered at the origin?

**Exercise 13.70.** Find the formulae for the curls of  $F + G$ ,  $gF$ .

**Exercise 13.71.** Prove that  $\operatorname{curl}(\operatorname{grad} f) = \vec{0}$ . Moreover, compute  $\operatorname{curl}(\operatorname{curl} F)$ .

**Exercise 13.72.** The electric field  $E$  induced by a changing magnetic field is also changing and follows *Ampere's law*

$$\int_C B \cdot d\vec{x} = \mu_0 \int_S J \cdot \vec{n} dA + \epsilon_0 \mu_0 \frac{d}{dt} \int_S E \cdot \vec{n} dA,$$

where  $J$  is the current density, and  $\mu_0$ ,  $\epsilon_0$  are some physical constants. Derive the differential version of Ampere's law, which is the other Maxwell's equation for electromagnetic fields.

## 2-Dimensional Fundamental Theorem of Calculus in $\mathbb{R}^n$

The key to derive Stokes' Theorem is to translate the integral along a surface in  $\mathbb{R}^3$  to the integral on a 2-dimensional body in  $\mathbb{R}^2$ . The argument certainly works in  $\mathbb{R}^n$ .

Let  $S$  be an oriented surface given by an orientation compatible regular parameterization  $\sigma(u, v): U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^n$ . For  $n > 3$ , the surface no longer has normal vectors that we can use to describe the orientation. Instead, the orientation is given by the parameterization itself. If the surface is covered by several parameterizations, then the derivatives of the transition maps are required to have positive determinants.

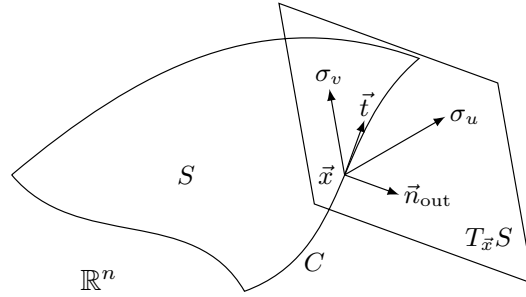
The boundary  $C$  of  $S$  corresponds to the boundary  $D$  of  $U$ . The boundary  $D$  is oriented in such a way that  $U$  is on the left side of  $D$ . The orientation of  $D$  gives the orientation of  $C$ .

The compatible orientation of  $C$  can also be described by using the outward normal vector  $\vec{n}_{\text{out}}$  along  $C$ . The orientation of  $C$  is represented by the tangent



vector  $\vec{t}$ . Both  $\{\vec{n}_{\text{out}}, \vec{t}\}$  and  $\{\sigma_u, \sigma_v\}$  are bases of the tangent space of  $S$  at boundary point. We require the direction of  $\vec{t}$  to be chosen in such a way that the matrix between the two bases has positive determinant. In terms of the exterior product, this means

$$\sigma_u \wedge \sigma_v = \lambda \vec{n} \wedge \vec{t}, \quad \lambda > 0.$$



**Figure 13.5.2.** *Compatible orientation of the boundary curve of a surface.*

For a continuously differentiable  $f$ , we have

$$\begin{aligned} \int_C f dx_j &= \int_D f \left( \frac{\partial x_j}{\partial u} du + \frac{\partial x_j}{\partial v} dv \right) \\ &= \int_U \left( \frac{\partial}{\partial u} \left( f \frac{\partial x_j}{\partial v} \right) - \frac{\partial}{\partial v} \left( f \frac{\partial x_j}{\partial u} \right) \right) dudv \\ &= \int_U \left( \frac{\partial f}{\partial u} \frac{\partial x_j}{\partial v} - \frac{\partial f}{\partial v} \frac{\partial x_j}{\partial u} \right) dudv \\ &= \int_U \left( \left( \sum_i \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial u} \right) \frac{\partial x_j}{\partial v} - \left( \sum_i \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial v} \right) \frac{\partial x_j}{\partial u} \right) dudv \\ &= \int_U \sum_i \frac{\partial f}{\partial x_i} \left( \frac{\partial x_i}{\partial u} \frac{\partial x_j}{\partial v} - \frac{\partial x_i}{\partial v} \frac{\partial x_j}{\partial u} \right) dudv \\ &= \int_U \sum_{i \neq j} \frac{\partial f}{\partial x_i} \det \frac{\partial(x_i, x_j)}{\partial(u, v)} dudv \\ &= \int_S \sum_{i \neq j} \frac{\partial f}{\partial x_i} dx_i \wedge dx_j. \end{aligned}$$

This leads to the following formula.

**Theorem 13.5.2 (Stokes' Theorem).** *Suppose  $S$  is an oriented surface in  $\mathbb{R}^n$  with compatible oriented boundary curve  $C$ . Then for any continuously differentiable  $f_1, \dots, f_n$  along the surface, we have*

$$\int_C \sum_j f_j dx_j = \int_S \sum_{i < j} \left( \frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} \right) dx_i \wedge dx_j.$$

### Potential: Antiderivative on $\mathbb{R}^n$

With the help of Stokes' Theorem in  $\mathbb{R}^n$ , Theorem 13.4.2 may be extended, with the same proof.

**Theorem 13.5.3.** *Suppose  $F = (f_1, \dots, f_n)$  is a continuous vector field on an open subset  $U \subset \mathbb{R}^n$ . Then the following are equivalent.*

1. *The integral  $\int_C F \cdot d\vec{x}$  along an oriented rectifiable curve  $C$  in  $U$  depends only on the beginning and end points of  $C$ .*
2. *There is a differentiable function  $\varphi$  on  $U$ , such that  $\nabla\varphi = F$ .*

Moreover, if any closed curve in  $U$  is the boundary of an orientable surface in  $U$ , and  $F$  is continuously differentiable, then the above is also equivalent to

3.  $\frac{\partial f_j}{\partial x_i} = \frac{\partial f_i}{\partial x_j}$  for all  $i, j$ .

The function  $\varphi$  in the theorem is the *potential* (or antiderivative) of the vector field  $F$  or the 1-form  $\omega = f_1 dx_1 + \dots + f_n dx_n = F \cdot d\vec{x}$ , and may be given by

$$\varphi(\vec{x}) = \int_{\vec{x}_0}^{\vec{x}} F \cdot d\vec{x}.$$

We note that  $\nabla\varphi = F$  is the same as  $d\varphi = F \cdot d\vec{x}$ .

A closed curve  $C$  is *homologous to 0* in  $U$  if it is the boundary of an oriented surface in  $U$ . Two oriented curves  $C_1$  and  $C_2$  in  $U$  with the same beginning and end points are *homologous* in  $U$  if the closed curve  $C_1 \cap (-C_2)$  is homologous to 0. Stokes' Theorem implies that, if the third statement is satisfied, then the integral of the 1-form along homologous curves are equal. The extra condition in Theorem 13.5.3 says that any two oriented curves with the same beginning and end points are homologous, so that the third statement implies the first.

A special case of the extra condition is that the subset  $U$  is *simply connected*. This means that any continuous map  $S^1 \rightarrow U$  extends to a continuous map  $B^2 \rightarrow U$ , so that we can choose the surface to be the disk in the extra condition.

**Example 13.5.3.** The vector field  $(f, g, h) = (yz(2x + y + z), zx(x + 2y + z), xy(x + y + 2z))$  satisfies

$$f_y = g_x = (2x + 2y + z)z, \quad h_y = g_z = (x + 2y + 2z)x, \quad f_z = h_x = (2x + y + 2z)y,$$

on the whole  $\mathbb{R}^3$ . Therefore the vector field has a potential, which can be computed by integrating along successive straight lines connecting  $(0, 0, 0)$ ,  $(x, 0, 0)$ ,  $(x, y, 0)$ ,  $(x, y, z)$ . The integral is zero on the first two segments, so that

$$\varphi(x, y, z) = \int_0^z xy(x + y + 2z)dz = xyz(x + y + z).$$

**Example 13.5.4.** The 1-form  $x_1 \cdots x_n \left( \frac{dx_1}{x_1} + \cdots + \frac{dx_n}{x_n} \right)$  satisfies

$$f_i = x_1 \cdots \hat{x}_i \cdots x_n, \quad \frac{\partial f_j}{\partial x_i} = x_1 \cdots \hat{x}_i \cdots \hat{x}_j \cdots x_n$$

on the whole  $\mathbb{R}^n$ . Therefore the 1-form has a potential. The potential function can be obtained by solving the system of equations

$$\frac{\partial \varphi}{\partial x_i} = x_1 \cdots \hat{x}_i \cdots x_n, \quad i = 1, \dots, n.$$

The solution of the first equation is

$$\varphi = x_1 \cdots x_n + \psi(x_2, \dots, x_n).$$

Substituting into the other equations, we get

$$\frac{\partial \varphi}{\partial x_i} = x_1 \cdots \hat{x}_i \cdots x_n + \frac{\partial \psi}{\partial x_i} = x_1 \cdots \hat{x}_i \cdots x_n, \quad i = 2, \dots, n.$$

Thus we conclude that  $\psi$  is a constant, and  $x_1 \cdots x_n$  is a potential function of the 1-form.

Solving the system of equations is actually the same as integrating along straight lines in coordinate directions like Example 13.5.3. Alternatively, we may also use the line integral to compute the potential function. We have already verified that the line integral depends only on the end points, and we may choose to integrate along the straight line  $\gamma(t) = t\vec{x}$  connecting  $\vec{0}$  to  $\vec{x}$

$$\begin{aligned} \varphi(\vec{x}) &= \int_0^1 \sum_{i=1}^n (tx_1) \cdots (\widehat{tx_i}) \cdots (tx_n) \frac{(tx_i)}{dt} dt \\ &= \int_0^1 \sum_{i=1}^n t^{n-1} x_1 \cdots x_n dt = x_1 \cdots x_n \int_0^1 nt^{n-1} dt = x_1 \cdots x_n. \end{aligned}$$

**Exercise 13.73.** Determine whether the integral of vector field or the 1-form is independent of the choice of the curves. In the independent case, find the potential function.

1.  $e^x(\cos yz, -z \sin yz, -y \sin yz)$ .
2.  $y^2 z^3 dx + 2xyz^3 dy + 2xyz^2 dz$ .
3.  $(y+z)dx + (z+x)dy + (x+y)dz$ .
4.  $(x_2, x_3, \dots, x_n, x_1)$ .
5.  $(x_1^2, \dots, x_n^2)$ .
6.  $x_1 x_2 \cdots x_n (x_1^{-1}, \dots, x_n^{-1})$ .

**Exercise 13.74.** Explain that the potential function is unique up to adding constants.

**Exercise 13.75.** In  $\mathbb{R}^3$ , explain that Theorem 13.5.3 tells us that, if all closed curves in  $U$  are homologous to 0, then  $F = \text{grad} \varphi$  for some  $\varphi$  on  $U$  if and only if  $\text{curl} F = \vec{0}$  on  $U$ .

**Exercise 13.76.** Suppose  $\vec{a}$  is a nonzero vector. Find condition on a function  $f(\vec{x})$  so that the vector field  $f(\vec{x})\vec{a}$  has a potential function.

**Exercise 13.77.** Find condition on a function  $f(\vec{x})$  so that the vector field  $f(\vec{x})\vec{x}$  has a potential function.

Exercise 13.78. Study the potential function of  $\frac{(y-z)dx + (z-x)dy + (x-y)dz}{(x-y)^2 + (y-z)^2}$ .

Exercise 13.79. Suppose a continuously second order differentiable function  $g(\vec{x})$  satisfies  $\nabla g(\vec{x}_0) \neq \vec{0}$ . Suppose  $f(\vec{x})$  is continuously differentiable near  $\vec{x}_0$ . Prove that the differential form

$$fdg = fg_{x_1}dx_1 + fg_{x_2}dx_2 + \cdots + fg_{x_n}dx_n$$

has a potential near  $\vec{x}_0$  if and only if  $f(\vec{x}) = h(g(\vec{x}))$  for a continuously differentiable  $h(t)$ .

## 13.6 Gauss' Theorem

Green's Theorem is the Fundamental Theorem of Calculus for 2-dimensional body in  $\mathbb{R}^2$ . The theorem can be extended to  $n$ -dimensional body in  $\mathbb{R}^n$ .

### 3-Dimensional Fundamental Theorem of Calculus in $\mathbb{R}^3$

The key step in the proof of Green's Theorem is to verify the case that  $U$  is the region between two functions. In  $\mathbb{R}^3$ , consider the 3-dimensional region between continuously differentiable functions  $h(x, y)$  and  $k(x, y)$  on  $V \subset \mathbb{R}^2$

$$U = \{(x, y, z): (x, y) \in V, k(x, y) \leq z \leq h(x, y)\}.$$

The boundary surface  $S$  of  $U$  is oriented to be compatible with the outward normal vector  $\vec{n}_{\text{out}}$ .

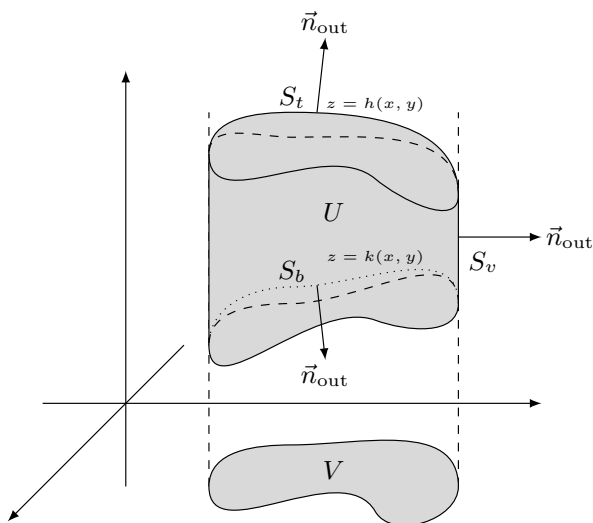


Figure 13.6.1. Gauss' Theorem for a special case.

We have

$$\begin{aligned}\int_U f_z dx dy dz &= \int_V \left( \int_{k(x,y)}^{h(x,y)} f_z(x, y, z) dz \right) dx dy && \text{(Fubini Theorem)} \\ &= \int_V (f(x, y, h(x, y)) - f(x, y, k(x, y))) dx dy. && \text{(Fundamental Th.)}\end{aligned}$$

On the other hand, the boundary surface  $S$  of  $U$  consists of three pieces. The top piece  $S_t$  is usually parameterized by

$$\sigma(x, y) = (x, y, h(x, y)), \quad (x, y) \in V.$$

By the computation in Example 13.2.6, the parameterization is compatible with the normal vector

$$\vec{n} = \frac{(-f_x, -f_y, 1)}{\sqrt{f_x^2 + f_y^2 + 1}}.$$

Since  $\vec{n}$  points upwards (the third coordinate is positive) and therefore outwards, we get  $\vec{n} = \vec{n}_{\text{out}}$ . Therefore the parameterization is compatible with the orientation of the top piece, and we have

$$\int_{S_t} f dx \wedge dy = \int_V f(\sigma(x, y)) dx dy = \int_V f(x, y, h(x, y)) dx dy.$$

The bottom piece  $S_b$  is usually parameterized by

$$\sigma(x, y) = (x, y, k(x, y)), \quad (x, y) \in V.$$

The parameterization is also compatible with the upward normal vector. Since the outward normal vector of the bottom piece points downward (the third coordinate is negative), we get  $\vec{n} = -\vec{n}_{\text{out}}$  for  $S_b$ , so that

$$\int_{S_b} f dx \wedge dy = - \int_V f(x, y, k(x, y)) dx dy.$$

We also have the usual parameterization of the side piece  $S_v$

$$\sigma(t, z) = (x(t), y(t), z), \quad k(x, y) \leq z \leq h(x, y),$$

where  $(x(t), y(t))$  is a parameterization of the boundary curve of  $V$ . Then

$$\int_{S_v} f dx \wedge dy = \pm \int_{a \leq t \leq b, k(x,y) \leq z \leq h(x,y)} f(x(t), y(t), z) \det \frac{\partial(x, y)}{\partial(t, z)} dt dz = 0.$$

We conclude that

$$\begin{aligned}\int_S f dx \wedge dy &= \left( \int_{S_t} + \int_{S_b} + \int_{S_v} \right) f dx \wedge dy \\ &= \int_V f(x, y, h(x, y)) dx dy - \int_V f(x, y, k(x, y)) dx dy \\ &= \int_U f_z dx dy dz.\end{aligned}$$

For a more general region  $U \subset \mathbb{R}^3$ , we may use the idea of the proof of Green's Theorem. In other words, we divide  $U$  into several special regions between graphs of functions and add the equality above for each region to get the same equality for  $U$ . Similar equalities for  $\int_S g dy \wedge dz$  and  $\int_S h dz \wedge dx$  can be established by rotating  $x$ ,  $y$  and  $z$ .

**Theorem 13.6.1 (Gauss' Theorem).** *Suppose  $U \subset \mathbb{R}^3$  is a region with the boundary surface  $S$  compatibly oriented with respect to the outward normal vector. Then for any continuously differentiable  $f$ ,  $g$ ,  $h$ , we have*

$$\int_S f dy \wedge dz + g dz \wedge dx + h dx \wedge dy = \int_U (f_x + g_y + h_z) dx dy dz.$$

Define the *divergence* of a vector field  $F = (f, g, h)$  to be

$$\operatorname{div} F = \nabla \cdot F = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} + \frac{\partial h}{\partial z}.$$

Then Gauss' Theorem means that the outward flux of a flow  $F = (f, g, h)$  is equal to the integral of the divergence of the flow on the solid

$$\int_S F \cdot \vec{n} dA = \int_U \operatorname{div} F dV.$$

**Example 13.6.1.** The volume of the region enclosed by a surface  $S \subset \mathbb{R}^3$  without boundary is

$$\int_S x dy \wedge dz = \int_S y dz \wedge dx = \int_S z dx \wedge dy = \frac{1}{3} \int_S (x, y, z) \cdot \vec{n} dA.$$

**Example 13.6.2.** In Example 13.2.9, the outgoing flux of the flow  $F = (x^2, y^2, z^2)$  through the ellipse  $\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} + \frac{(z-z_0)^2}{c^2} = 1$  is computed by surface integral. Alternatively, by Gauss' theorem, the flux is

$$\begin{aligned} \int F \cdot \vec{n} dA &= \int_{\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} + \frac{(z-z_0)^2}{c^2} \leq 1} 2(x+y+z) dx dy dz \\ &= \int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1} 2(x+x_0+y+y_0+z+z_0) dx dy dz. \end{aligned}$$

By the transform  $\vec{x} \rightarrow -\vec{x}$ , we get

$$\int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1} (x+y+z) dx dy dz = 0.$$

Therefore the flux is

$$\int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1} 2(x_0+y_0+z_0) dx dy dz = \frac{8\pi R^3}{3} abc(x_0+y_0+z_0).$$

**Example 13.6.3.** To compute the upward flux of  $F = (xz, -yz, (x^2 + y^2)z)$  through the surface  $S$  given by  $0 \leq z = 4 - x^2 - y^2$ , we introduce the disk  $D = \{(x, y, 0) : x^2 + y^2 \leq 4\}$  on the  $(x, y)$ -plane. Taking the normal direction of  $D$  to be  $(0, 0, 1)$ , the surface  $S \cup (-D)$  is the boundary of the region  $U$  given by  $0 \leq z \leq 4 - x^2 - y^2$ . By Gauss' theorem, the flux through  $S$  is

$$\begin{aligned} \int_S F \cdot \vec{n} dA &= \int_{S \cup (-D)} F \cdot \vec{n} dA + \int_D F \cdot \vec{n} dA \\ &= \int_U (z - z + x^2 + y^2) dx dy dz + \int_{x^2 + y^2 \leq 4} F(x, y, 0) \cdot (0, 0, 1) dA \\ &= \int_U (x^2 + y^2) dx dy dz = \int_{0 \leq r \leq 2, 0 \leq \theta \leq 2\pi} r^2(4 - r^2) r dr d\theta = \frac{32\pi}{3}. \end{aligned}$$

**Example 13.6.4.** The gravitational field created by a mass  $M$  at point  $\vec{x}_0 \in \mathbb{R}$  is

$$G = -\frac{M}{\|\vec{x} - \vec{x}_0\|_2^3}(\vec{x} - \vec{x}_0).$$

A straightforward computation shows  $\text{div} G = 0$ . Suppose  $U$  is a region with compatibly oriented boundary surface  $S$  and  $\vec{x}_0 \notin S$ . If  $\vec{x}_0 \notin U$ , then by Gauss' theorem, the outward flux  $\int_S G \cdot \vec{n} dA = 0$ . If  $\vec{x}_0 \in U$ , then let  $B_\epsilon$  be the ball of radius  $\epsilon$  centered at  $\vec{x}_0$ . The boundary of the ball is the sphere  $S_\epsilon$ , which we give an orientation compatible with the outward normal vector  $\vec{n} = \frac{\vec{x} - \vec{x}_0}{\|\vec{x} - \vec{x}_0\|_2}$ . For sufficiently small  $\epsilon$ , the ball is contained in  $U$ . Moreover,  $U - B_\epsilon$  is a region not containing  $\vec{x}_0$  and has compatibly oriented boundary surface  $S \cup (-S_\epsilon)$ . Therefore

$$\int_S G \cdot \vec{n} dA = \int_{S_\epsilon} G \cdot \vec{n} dA = \int_{\|\vec{x} - \vec{x}_0\|_2 = \epsilon} -\frac{M}{\epsilon^3}(\vec{x} - \vec{x}_0) \cdot \frac{\vec{x} - \vec{x}_0}{\epsilon} dA = -\frac{M}{\epsilon^2} \int_{S_\epsilon} dA = -4\pi M.$$

More generally, the gravitational field created by several masses at various locations is the sum of the individual gravitational field. The outward flux of the field through  $S$  is then  $-4\pi$  multiplied to the total mass contained in  $A$ . In particular, the flux is independent of the specific location of the mass, but only on whether the mass is inside  $A$  or not. This is called *Gauss' Law*.

**Exercise 13.80.** Compute the flux.

1. Outward flux of  $(x^3, x^2y, x^2z)$  through boundary of the solid  $x^2 + y^2 \leq a^2, 0 \leq z \leq b$ .
2. Inward flux of  $(xy^2, yz^2, zx^2)$  through the ellipse  $\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} + \frac{(z - z_0)^2}{c^2} = 1$ .
3. Upward flux of  $(x^3, y^3, z^3)$  through the surface  $z = x^2 + y^2 \leq 1$ .
4. Outward flux of  $(x^2, y^2, -2(x + y)z)$  through the torus in Example 8.1.17.

**Exercise 13.81.** Suppose  $U \subset \mathbb{R}^3$  is a convex region with boundary surface  $S$ . Suppose  $\vec{a}$  is a vector in the interior of  $U$ , and  $p$  is the distance from  $\vec{a}$  to the tangent plane of  $S$ . Compute  $\int_S p dA$ . Moreover, for the special case  $S$  is the ellipse  $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$  and  $\vec{a} = (0, 0, 0)$ , compute  $\int_S \frac{1}{p} dA$ .

**Exercise 13.82.** Find the formulae for the divergences of  $F + G$ ,  $gF$ ,  $F \times G$ .

**Exercise 13.83.** Prove that  $\operatorname{div}(\operatorname{curl} F) = 0$ . Moreover, compute  $\operatorname{div}(\operatorname{grad} f)$  and  $\operatorname{grad}(\operatorname{div} F)$ .

**Exercise 13.84.** Find all the functions  $f$  on  $\mathbb{R}^3$ , such that the integral of the differential form  $f dy \wedge dz + z dx \wedge dy$  on any sphere is 0.

### **$n$ -Dimensional Fundamental Theorem of Calculus in $\mathbb{R}^n$**

Such a fundamental theorem is inevitable, and the proof is routine. First we consider the special region

$$U = \{(\vec{u}, x_n) : \vec{u} \in V, k(\vec{u}) \leq x_n \leq h(\vec{u})\}, \quad \vec{u} = (x_1, \dots, x_{n-1}).$$

The boundary  $S$  of  $U$  is oriented to be compatible with the outward normal vector  $\vec{n}_{\text{out}}$ .

By Fubini Theorem and the classical Fundamental Theorem of Calculus, we have

$$\int_U f_{x_n} dx_1 \cdots dx_n = \int_V (f(\vec{u}, h(\vec{u})) - f(\vec{u}, k(\vec{u}))) dx_1 \cdots dx_{n-1}.$$

By the computation in Example 13.3.3, the usual parameterizations of the top piece  $S_t$

$$\sigma(\vec{u}) = (\vec{u}, h(\vec{u})), \quad \vec{u} \in V,$$

and the bottom piece  $S_b$

$$\sigma(\vec{u}) = (\vec{u}, k(\vec{u})), \quad \vec{u} \in V,$$

are compatible with the normal vector with last coordinate having sign  $(-1)^{n-1}$ . Since the  $\vec{n}_{\text{out}}$  has positive last coordinate on  $S_t$  and negative last coordinate on  $S_b$ , we get

$$\int_{S_t} f dx_1 \wedge \cdots \wedge dx_{n-1} = (-1)^{n-1} \int_V f(\vec{u}, h(\vec{u})) dx_1 \cdots dx_{n-1}, \quad (13.6.1)$$

$$\int_{S_b} f dx_1 \wedge \cdots \wedge dx_{n-1} = (-1)^n \int_V f(\vec{u}, k(\vec{u})) dx_1 \cdots dx_{n-1}. \quad (13.6.2)$$

Similar to the earlier discussions, we also have

$$\int_{S_v} f dx_1 \wedge \cdots \wedge dx_{n-1} = 0.$$

Then we get

$$\int_S (-1)^{n-1} f dx_1 \wedge \cdots \wedge dx_{n-1} = \int_U f_{x_n} dx_1 \cdots dx_n.$$

By the same argument, especially using the computations in Example 13.3.3 and Exercise 13.49, we get

$$\int_S (-1)^{i-1} f dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n = \int_U f_{x_i} dx_1 \cdots dx_n.$$



**Theorem 13.6.2 (Gauss' Theorem).** *Suppose  $U \subset \mathbb{R}^n$  is a region with the boundary submanifold  $S$  compatibly oriented with respect to the outward normal vector. Then for any continuously differentiable  $f_1, \dots, f_n$ , we have*

$$\int_S \sum (-1)^{i-1} f_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n = \int_U \left( \frac{\partial f_1}{\partial x_1} + \cdots + \frac{\partial f_n}{\partial x_n} \right) dx_1 \cdots dx_n.$$

If we define the divergence of a vector field on  $\mathbb{R}^n$  to be

$$\operatorname{div} F = \frac{\partial f_1}{\partial x_1} + \cdots + \frac{\partial f_n}{\partial x_n}.$$

then Gauss' Theorem can be rewritten as

$$\int_S F \cdot \vec{n} dV = \int_U \operatorname{div} F d\mu_{\vec{x}}.$$

We obtained Stokes' Theorems 13.5.1 and 13.5.2 by “transferring” Green's Theorem to a surface in  $\mathbb{R}^n$ . By the same method, we can “transfer” Gauss' Theorems 13.6.1 and 13.6.2 to submanifolds of higher dimensional Euclidean spaces. However, we will leave the discussion to the next chapter, because there is another deep insight that the high dimensional Fundamental Theorems of Calculus does not require Euclidean space at all!

## 13.7 Additional Exercise

### Gauss Map

Let  $M$  be an oriented submanifold of  $\mathbb{R}^n$  of dimension  $n - 1$ . The normal vector can be considered as a map  $\nu = \vec{n}: M \rightarrow S^{n-1} \subset \mathbb{R}^n$ , called the *Gauss map*.

**Exercise 13.85.** Prove that the derivative  $\nu'$  maps  $T_{\vec{x}}M$  to the subspace  $T_{\vec{x}}M \subset \mathbb{R}^n$ . The map  $\nu': T_{\vec{x}}M \rightarrow T_{\vec{x}}M$  is called the *Weingarten map*.

**Exercise 13.86.** Prove that the Weingarten map is self-adjoint:  $\nu'(\vec{u}) \cdot \vec{v} = \vec{u} \cdot \nu'(\vec{v})$  for any  $\vec{u}, \vec{v} \in T_{\vec{x}}M$ .

**Exercise 13.87.** Compute the Weingarten map.

1.  $M = \{\vec{x}: \vec{a} \cdot \vec{x} = c\}$  is a hyperplane of  $\mathbb{R}^n$ .
2.  $M = S_R^{n-1}$  is the sphere of radius  $R$  in  $\mathbb{R}^n$ .
3.  $M = S_R^{n-2} \times \mathbb{R}$  is the cylinder of radius  $R$  in  $\mathbb{R}^n$ .

### Principal Curvatures

For an oriented submanifold  $M$  of  $\mathbb{R}^n$  of dimension  $n - 1$ , its thickening map

$$\phi(\vec{x}, t) = \vec{x} + t\vec{n}(\vec{x}): M \times [a, b] \rightarrow \mathbb{R}^n,$$

is injective for small  $a, b$  and describes an open neighborhood  $N_{[a,b]}$  of  $M$  in  $\mathbb{R}^n$  in case  $a < 0 < b$ .

Let  $\sigma: U \rightarrow M$  is an orientation compatible parameterization of  $M$ . Let  $W$  be the matrix of the Weingarten map with respect to the basis  $\sigma_{u_1}, \dots, \sigma_{u_{n-1}}$ . Then

$$\det(I + tW) = 1 + t(n-1)H_1 + \dots + t^k \frac{(n-1)!}{(k-1)!(n-k)!} H_{k-1} + \dots + t^{n-1} H_{n-1}.$$

We call  $H_1$  the *mean curvature* and call  $H_{n-1}$  the *Gaussian curvature*.

**Exercise 13.88.** Explain that  $H_k$  are independent of the choice of the parameterization  $\sigma$ .

**Exercise 13.89.** Prove that  $\|\phi_{u_1} \wedge \dots \wedge \phi_{u_{n-1}} \wedge \phi_t\|_2 = \det(I + tL) \|\sigma_{u_1} \wedge \dots \wedge \sigma_{u_{n-1}}\|_2$  for small  $t$ .

**Exercise 13.90.** Prove that the volume of  $N_{[0,b]}$  is

$$b \text{Vol}(M) + \frac{n-1}{2} b^2 \int_M H_1 dV + \dots + \frac{(n-1)!}{k!(n-k)!} b^k \int_M H_k dV + \dots + \frac{1}{n} b^n \int_M H_{n-1} dV.$$

What about the volumes of  $N_{[a,0]}$  and  $N_{[a,b]}$ ?

**Exercise 13.91.** Derive the formula in Example 13.3.1.

## Chapter 14

# Manifold

## 14.1 Manifold

In Section 8.4, we defined an *n-dimensional submanifold* of  $\mathbb{R}^N$  to be a subset  $M \subset \mathbb{R}^N$ , such that near any point in  $M$ , the subset is the graph of a continuously differentiable map  $f$  of some choice of  $N - n$  coordinates  $\vec{z}$  in terms of the other  $n$  coordinates  $\vec{y}$ . Specifically, if vectors of  $\mathbb{R}^N$  are written as  $\vec{x} = (\vec{y}, \vec{z})$ ,  $\vec{y} \in \mathbb{R}^n$ ,  $\vec{z} \in \mathbb{R}^{N-n}$ , then

$$\varphi(\vec{y}) = (\vec{y}, f(\vec{y})): U \rightarrow M$$

is a regular parameterization near the point, in the sense that the partial derivative of  $\varphi$  in  $\vec{y}$  is invertible. Conversely, by Proposition 8.4.2, a submanifold of  $\mathbb{R}^N$  is a subset  $M$ , such that there is a regular parameterization near any point of  $M$ .

Another motivation for differentiable manifolds is from the discussion leading to the definition (??) of the integral of a form on an oriented surface. The discussion suggests that the integral can be defined as long as there are orientation compatible parameterizations of pieces of the surface. The key observation is that the ambient Euclidean space  $\mathbb{R}^N$  is not needed in the definition of the integral.

### Differentiable Manifold

The following is the preliminary definition of differentiable manifolds. The rigorous definition requires extra topological conditions and will be given in Definition 14.2.10 after we know more about the topology of manifolds.

**Definition 14.1.1 (Preliminary).** A *differentiable manifold* of dimension  $n$  is a set  $M$  and a collection of injective maps

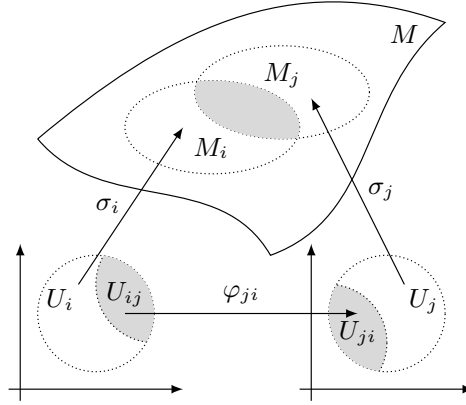
$$\sigma_i: U_i \rightarrow M_i = \sigma_i(U_i) \subset M,$$

such that the following are satisfied.

1.  $M$  is covered by  $M_i$ :  $M = \cup M_i$ .
2.  $U_i$  are open subsets of  $\mathbb{R}^n$ .
3.  $U_{ij} = \sigma_i^{-1}(M_i \cap M_j)$  are open subsets of  $\mathbb{R}^n$ .
4. The *transition maps*  $\varphi_{ji} = \sigma_j^{-1} \circ \sigma_i: U_{ij} \rightarrow U_{ji}$  are continuously differentiable.

If the transition maps are  $r$ -th order continuously differentiable, then we say  $M$  is an *r-th order differentiable manifold*, or  *$C^r$ -manifold*. If the transition maps are continuously differentiable of any order, then  $M$  is a *smooth manifold*, or  *$C^\infty$ -manifold*. On the other hand, if the transition maps are only continuous, then  $M$  is a *topological manifold*, or  *$C^0$ -manifold*.

The map  $\sigma_i$  is a (*coordinate*) *chart*. A collection  $\{\sigma_i\}$  of coordinate charts satisfying the conditions of the definition is an *atlas* of the manifold. The same manifold may allow different choices of atlases. For example, any way of expressing the circle as a union of some open circular intervals gives an atlas of the circle. On the other hand, the concept of manifold should be independent of the choice



**Figure 14.1.1.** Transition map between overlapping parameterizations.

of atlas. To solve the problem of dependence on the atlas, we say an injection  $\sigma: U \rightarrow \sigma(U) \subset M$  from an open subset  $U \subset \mathbb{R}^n$  is a (*compatible*) *chart* if for each existing chart  $\sigma_i$ ,  $\sigma_i^{-1}(M_i \cap \sigma(U))$  is an open subset of  $U_i$ , and the map

$$\sigma^{-1} \circ \sigma_i: \sigma_i^{-1}(M_i \cap \sigma(U)) \rightarrow U$$

is continuously differentiable. A unique atlas of a manifold can then be constructed by including all the compatible charts to form the maximal atlas.

**Example 14.1.1.** Any open subset  $U \subset \mathbb{R}^n$  is a manifold. We can simply use the identity map  $U \rightarrow U$  as an atlas. Any other chart is simply an injective continuously differentiable map  $\sigma: V \subset \mathbb{R}^n \rightarrow U$  from an open subset  $V \subset \mathbb{R}^n$ , such that  $\sigma(V)$  is open, and  $\sigma^{-1}: \sigma(V) \rightarrow V$  is also continuously differentiable.

**Example 14.1.2.** In Example 8.4.4, the unit circle  $S^1 = \{(x, y): x^2 + y^2 = 1\}$  has regular parameterizations given by expressing one coordinate as the function of the other coordinate

$$\begin{aligned}\sigma_1(u) &= (u, \sqrt{1-u^2}): U_1 = (-1, 1) \rightarrow M_1 \subset S^1, \\ \sigma_{-1}(u) &= (u, -\sqrt{1-u^2}): U_{-1} = (-1, 1) \rightarrow M_{-1} \subset S^1, \\ \sigma_2(u) &= (\sqrt{1-u^2}, u): U_2 = (-1, 1) \rightarrow M_2 \subset S^1, \\ \sigma_{-2}(u) &= (-\sqrt{1-u^2}, u): U_{-2} = (-1, 1) \rightarrow M_{-2} \subset S^1.\end{aligned}$$

Here  $M_1, M_{-1}, M_2, M_{-2}$  are respectively the upper, lower, right, and left half circles. The overlapping  $M_1 \cap M_2$  is the upper right quarter of the circle. The transition map  $\varphi_{21}(u) = \sigma_2^{-1} \circ \sigma_1(u) = w$  is obtained by solving  $\sigma_1(u) = \sigma_2(w)$ , or  $(*, \sqrt{1-u^2}) = (*, w)$ . Therefore

$$\varphi_{21}(u) = \sqrt{1-u^2}: U_{12} = (0, 1) \rightarrow U_{21} = (0, 1).$$

We can similarly get the other three transition maps, such as

$$\varphi_{(-2)1}(u) = \sqrt{1-u^2}: U_{1(-2)} = (-1, 0) \rightarrow U_{(-2)1} = (0, 1).$$

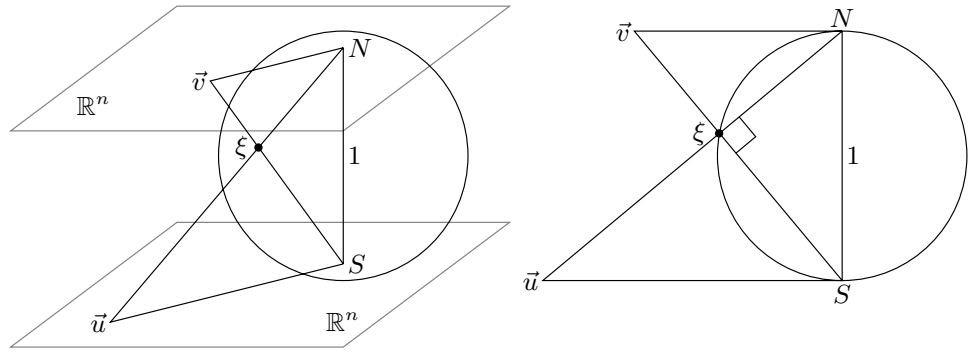
All transition maps are smooth.

**Example 14.1.3.** Consider the sphere  $S^n$  of radius  $\frac{1}{2}$  (or diameter 1). Fixing two antipodal points  $N$  and  $S$  as the north and south poles, we have the *stereographical projection* from the north pole

$$\sigma_N(\vec{u}) = \xi: \mathbb{R}^n \rightarrow S^n - \{N\}.$$

The chart misses the north pole. To cover the whole sphere, we also need to use the stereographical projection from the south pole

$$\sigma_S(\vec{v}) = \xi: \mathbb{R}^n \rightarrow S^n - \{S\}.$$



**Figure 14.1.2.** Stereographic projections.

The overlapping of the two charts is  $S^n - \{N, S\}$ , and

$$U_{NS} = \sigma_N^{-1}(S^n - \{N, S\}) = \mathbb{R}^n - \vec{0}, \quad U_{SN} = \sigma_S^{-1}(S^n - \{N, S\}) = \mathbb{R}^n - \vec{0}.$$

The transition map is

$$\varphi_{NS}(\vec{u}) = \sigma_N^{-1} \circ \sigma_S(\vec{u}) = \vec{v}: \mathbb{R}^n - \vec{0} \rightarrow \mathbb{R}^n - \vec{0}.$$

Note that by the obvious identification of the two  $\mathbb{R}^n$  (explicitly given by  $(x_1, \dots, x_n, \frac{1}{2}) \leftrightarrow (x_1, \dots, x_n, -\frac{1}{2})$ ),  $\vec{u}$  and  $\vec{v}$  point to the same direction, and their Euclidean lengths are related by  $\|\vec{u}\|_2 \|\vec{v}\|_2 = 1^2 = 1$ . Therefore we get

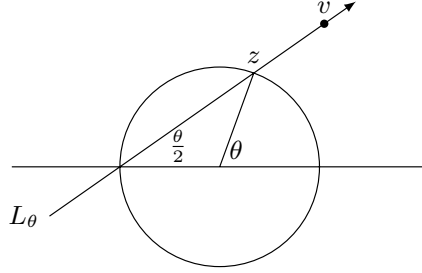
$$\varphi_{NS}(\vec{v}) = \frac{\vec{v}}{\|\vec{v}\|_2^2}.$$

The transition map is smooth.

**Example 14.1.4 (Open Möbius Band).** Let  $L_\theta$  be the line passing through the point  $z(\theta) = (\cos \theta, \sin \theta)$  on the unit circle at angle  $\theta$  and with slope  $\frac{1}{2}\theta$ . The *Möbius band* is

$$M = \{(z, v): z \in S^1, v \in L_\theta\}.$$

Although all  $L_\theta$  intersect at  $v = (-1, 0)$ , we consider  $(z, v)$  and  $(z', v)$  to be different points of  $M$  if  $z \neq z'$ . On the other hand, if  $z(\theta) = z(\theta')$ , then  $\theta - \theta'$  is an integer multiple of  $2\pi$ , and  $L_\theta$  and  $L_{\theta'}$  are the same line. So there is no ambiguity in the definition of  $M$ .



**Figure 14.1.3.** *Open Möbius band.*

Corresponding to the covering of the circle by two charts  $(0, 2\pi)$  and  $(-\pi, \pi)$ , we have two charts for the Möbius band, given by the same formula

$$\sigma_i(\theta, t) = ((\cos \theta, \sin \theta), (\cos \theta, \sin \theta) + t(\cos \tfrac{1}{2}\theta, \sin \tfrac{1}{2}\theta)),$$

but with different domain

$$\sigma_0: U_0 = (0, 2\pi) \times \mathbb{R} \rightarrow M, \quad \sigma_1: U_1 = (-\pi, \pi) \times \mathbb{R} \rightarrow M.$$

The transition map  $\varphi_{10}(\theta, t) = (\theta', t')$  on the overlapping  $(0, \pi) \times \mathbb{R} \sqcup (\pi, 2\pi) \times \mathbb{R}$  is obtained by solving  $\sigma_0(\theta, t) = \sigma_1(\theta', t')$ . The solution consists of two parts

$$\begin{aligned} \varphi_{10}^+(\theta, t) &= (\theta, t): (0, \pi) \times \mathbb{R} \rightarrow (0, \pi) \times \mathbb{R}, \\ \varphi_{10}^-(\theta, t) &= (\theta - 2\pi, -t): (\pi, 2\pi) \times \mathbb{R} \rightarrow (-\pi, 0) \times \mathbb{R}. \end{aligned}$$

Both are smooth maps.

**Example 14.1.5.** The *real projective space*  $\mathbb{R}P^n$  is the set of the lines in  $\mathbb{R}^{n+1}$  passing through the origin. Such a line is a 1-dimensional subspace, and is spanned by a nonzero vector  $(x_0, x_1, \dots, x_n)$ . Denoting the span by  $[x_0, x_1, \dots, x_n]$ , we have

$$\mathbb{R}P^n = \{[x_0, x_1, \dots, x_n]: \text{some } x_i \neq 0\}.$$

If  $x_i \neq 0$ , then  $(x_0, x_1, \dots, x_n)$  and  $\frac{1}{x_i}(x_0, x_1, \dots, x_n)$  span the same subspace. This leads to the injective maps

$$\sigma_i(x_1, \dots, x_n) = [x_1, \dots, x_i, 1, x_{i+1}, \dots, x_n]: \mathbb{R}^n \rightarrow \mathbb{R}P^n, \quad 0 \leq i \leq n,$$

with the images  $M_i = \{[x_0, x_1, \dots, x_n]: x_i \neq 0\}$ .

For the overlapping  $M_0 \cap M_1 = \{[x_0, x_1, \dots, x_n]: x_0 \neq 0 \text{ and } x_1 \neq 0\}$ , we have

$$U_{01} = U_{10} = \{(x_1, \dots, x_n): x_1 \neq 0\}.$$

The transition map  $\varphi_{10}(x_1, x_2, \dots, x_n) = (y_1, y_2, \dots, y_n)$  is obtained from

$$\sigma_0(x_1, \dots, x_n) = [1, x_1, x_2, \dots, x_n] = [y_1, 1, y_2, \dots, y_n] = \sigma_1(y_1, \dots, y_n).$$

Therefore

$$\varphi_{10}(x_1, x_2, \dots, x_n) = \left( \frac{1}{x_1}, \frac{x_2}{x_1}, \dots, \frac{x_n}{x_1} \right).$$

We can get similar formulae for the other transition maps. All the transition maps are smooth.

**Example 14.1.6.** The *complex projective space*  $\mathbb{C}P^n$  is the set of all complex 1-dimensional subspaces of  $\mathbb{C}^{n+1}$ . Denoting by  $[z_0, z_1, \dots, z_n]$  the complex subspace spanned by (nonzero) complex vector  $(z_0, z_1, \dots, z_n)$ , we have

$$\mathbb{C}P^n = \{[z_0, z_1, \dots, z_n] : \text{some } z_i \neq 0\}.$$

Like  $\mathbb{R}P^n$ , the complex projective space is also a manifold. We specifically look at  $\mathbb{C}P^1$ , which like the real case is covered by two charts ( $w$  is a complex number)

$$\begin{aligned}\sigma_0(w) &= [1, w] : U_0 = \mathbb{C} \rightarrow M_0 = \{[z_0, z_1] : z_0 \neq 0\}, \\ \sigma_1(w) &= [w, 1] : U_1 = \mathbb{C} \rightarrow M_1 = \{[z_0, z_1] : z_1 \neq 0\}.\end{aligned}$$

The overlapping  $M_0 \cap M_1$  consists of those  $[z_0, z_1]$  with both  $z_0$  and  $z_1$  nonzero. The transition map is

$$\varphi_{10}(w) = \frac{1}{w} : U_{01} = \mathbb{C} - 0 \rightarrow U_{10} = \mathbb{C} - 0.$$

In terms of  $w = u + iv$ ,  $u, v \in \mathbb{R}$ , the transition map is

$$\varphi_{10}(u, v) = \left( \frac{u}{u^2 + v^2}, \frac{-v}{u^2 + v^2} \right) : \mathbb{R}^2 - (0, 0) \rightarrow \mathbb{R}^2 - (0, 0).$$

**Exercise 14.1.** Show that the torus is a differentiable manifold by constructing an atlas.

**Exercise 14.2.** Show that the special linear group  $SL(2)$  in Example 8.4.9 is a differentiable manifold by constructing an atlas.

**Exercise 14.3.** The points on the circle can be described by angles  $\theta$ , with  $\theta$  and  $\theta + 2\pi$  corresponding to the same point. Find the atlas consisting of two circle intervals  $(-\delta, \pi + \delta)$  and  $(-\pi - \delta, \delta)$ , where  $0 < \delta < \pi$ . Find the transition map.

**Exercise 14.4.** Show that the two atlases of the circle in Example 14.1.2 and Exercises 14.3 are compatible and therefore give the same manifold.

**Exercise 14.5.** In Example 14.1.3, find the formula of  $\xi$  in terms of  $\vec{u}$  and  $\vec{v}$ .

**Exercise 14.6.** Extend Example 14.1.2 to the unit sphere  $S^n$  in  $\mathbb{R}^{n+1}$ . Then (after scaling by 2 to match radius  $\frac{1}{2}$ ) compare with the stereographical projection charts in Example 14.1.3.

**Exercise 14.7.** Describe the atlas of the circle given by the stereographical projection in Example 14.1.3. Then compare the atlas with the atlas of the real projective space  $\mathbb{R}P^1$  in Example 14.1.5. Explain why the circle and  $\mathbb{R}P^1$  are the same manifold.

**Exercise 14.8.** Show that the sphere  $S^2$  and the complex projective space  $\mathbb{C}P^1$  are the same by comparing the stereographic projection atlas of  $S^2$  with the atlas of  $\mathbb{C}P^1$  in Example 14.1.6.

**Exercise 14.9.** Describe an atlas for the complex projective space  $\mathbb{C}P^n$ .



**Exercise 14.10.** Show that if we change  $\frac{1}{2}\theta$  to  $\frac{5}{2}\theta$  in Example 14.1.4, then we still get the Möbius band.

**Exercise 14.11 (Klein Bottle).** The *Klein bottle* is obtained by identifying the boundaries of the square  $[-1, 1] \times [-1, 1]$  as follows

$$(x, 1) \sim (-x, -1), \quad (1, y) \sim (-1, y).$$

Show that the Klein bottle is a differentiable manifold.

**Exercise 14.12 (Mapping Torus).** Let  $L$  be an invertible linear transform on  $\mathbb{R}^n$ . Show that the *mapping torus* obtained by glueing two ends of  $\mathbb{R}^n \times [0, 1]$  by  $(\vec{x}, 1) \sim (L(\vec{x}), 0)$  is a differentiable manifold.

**Exercise 14.13.** Suppose  $M$  and  $N$  are manifolds of dimensions  $m$  and  $n$ . Prove that the product  $M \times N$  is a manifold of dimension  $m + n$ . Specifically, prove that if  $\{\sigma_i: U_i \rightarrow M_i\}$  and  $\{\tau_j: V_j \rightarrow N_j\}$  are atlases of  $M$  and  $N$ , then  $\{\sigma_i \times \tau_j: U_i \times V_j \rightarrow M_i \times N_j\}$  is an atlas of  $M \times N$ .

**Exercise 14.14.** Show that it is always possible to find an atlas of a differentiable manifold, such that every  $U_i = \mathbb{R}^n$ .

## Manifold with Boundary

Definitions 14.1.1 only gives *manifolds without boundary*. If we allow  $U_i$  and  $U_{ij}$  to be open subsets of either the whole Euclidean space  $\mathbb{R}^n$  or the half Euclidean space

$$\mathbb{R}_+^n = \{(u_1, u_2, \dots, u_{n-1}, u_n) : u_n \geq 0\},$$

then we get *manifolds with boundary*, and the *boundary*  $\partial M$  consists of those points corresponding to

$$\partial \mathbb{R}_+^n = \{(u_1, u_2, \dots, u_{n-1}, 0)\} = \mathbb{R}^{n-1} \times 0.$$

In other words,

$$\partial M = \cup_{U_i \subset \mathbb{R}_+^n} \sigma_i(U_i \cap \mathbb{R}^{n-1} \times 0).$$

In the extended definition, a subset of  $\mathbb{R}_+^n$  is open if it is of the form  $\mathbb{R}_+^n \cap U$  for some open subset  $U$  of  $\mathbb{R}^n$ . Moreover, a map on an open subset of  $\mathbb{R}_+^n$  is (continuously) differentiable if it is the restriction of a (continuously) differentiable map on an open subset of  $\mathbb{R}^n$ . This makes it possible for us to talk about the differentiability of the transition maps between open subsets of  $\mathbb{R}_+^n$ .

**Proposition 14.1.2.** *The boundary of an  $n$ -dimensional differentiable manifold is an  $(n - 1)$ -dimensional differentiable manifold without boundary.*

*Proof.* An atlas for the boundary  $\partial M$  can be obtained by restricting an atlas of  $M$  to  $\mathbb{R}^{n-1} \times 0$

$$\tau_i = \sigma_i|_{\mathbb{R}^{n-1} \times 0}: V_i = U_i \cap \mathbb{R}^{n-1} \times 0 \rightarrow (\partial M)_i = \sigma_i(U_i \cap \mathbb{R}^{n-1} \times 0).$$

The transition maps between the charts  $\tau_i$  are continuously differentiable because they are the restrictions of continuously differentiable transition maps of  $M$  (which are again restrictions of continuously differentiable maps on open subsets of  $\mathbb{R}^n$ ). Moreover,  $\partial M$  has no boundary because  $V_i$  are open subsets of the whole Euclidean space  $\mathbb{R}^{n-1}$  (the half space  $\mathbb{R}_+^{n-1}$  is not used for the atlas).  $\square$

**Example 14.1.7.** The interval  $M = [0, 1]$  is covered by charts

$$\begin{aligned}\sigma_1(t) &= t: U_1 = [0, 1) \rightarrow M_1 = [0, 1) \subset M, \\ \sigma_2(t) &= 1 - t: U_2 = [0, 1) \rightarrow M_2 = (0, 1] \subset M.\end{aligned}$$

Here  $U_1 = U_2 = [0, 1) = (-1, 1) \cap \mathbb{R}_+$  are open subsets of the half line  $\mathbb{R}_+ = [0, +\infty)$ , and the transition map  $\varphi_{21}(t) = 1 - t: (0, 1) \rightarrow (0, 1)$  is continuously differentiable. The boundary of the interval is

$$\partial[0, 1] = \sigma_1(0) \cup \sigma_2(0) = \{0, 1\}.$$

**Example 14.1.8.** The disk  $M = \{(x, y): x^2 + y^2 \leq 1\}$  is covered by charts

$$\begin{aligned}\sigma_0(x, y) &= (x, y): U_0 = \{(x, y): x^2 + y^2 < 1\} \rightarrow M_0 \subset M, \\ \sigma_1(r, \theta) &= (r \cos \theta, r \sin \theta): U_1 = (0, 1] \times (-\delta, \pi + \delta) \rightarrow M_1 \subset M, \\ \sigma_2(r, \theta) &= (r \cos \theta, r \sin \theta): U_2 = (0, 1] \times (-\pi - \delta, \delta) \rightarrow M_2 \subset M.\end{aligned}$$

Here we use Exercise 14.3 for the circular direction. Note that to make the formulae more manageable, we relaxed the choice of  $U_i$  to be open subsets of *any* half Euclidean space, and the half plane used here is  $(-\infty, 1] \times \mathbb{R}$ . If we insist on using the standard half space, then the charts have more cumbersome formulae

$$\begin{aligned}\tilde{\sigma}_0(x, y) &= (x, y): \tilde{U}_0 = \{(x, y): x^2 + y^2 < 1\} \rightarrow M_0 \subset M, \\ \tilde{\sigma}_1(\theta, t) &= ((1 - t) \cos \theta, (1 - t) \sin \theta): \tilde{U}_1 = (-\delta, \pi + \delta) \times [0, +\infty) \rightarrow M_1 \subset M, \\ \tilde{\sigma}_2(\theta, t) &= ((1 - t) \cos \theta, (1 - t) \sin \theta): \tilde{U}_2 = (-\pi - \delta, \delta) \times [0, +\infty) \rightarrow M_2 \subset M.\end{aligned}$$

We can easily find the transition maps

$$\begin{aligned}\varphi_{01}(x, y) &= (r \cos \theta, r \sin \theta): U_{10} = (0, 1) \times (-\delta, \pi + \delta) \\ &\rightarrow U_{01} = \{(x, y): 0 < x^2 + y^2 < 1, -\delta < \theta(x, y) < \pi + \delta\}, \\ \varphi_{02}(x, y) &= (r \cos \theta, r \sin \theta): U_{20} = (0, 1) \times (-\pi - \delta, \delta) \\ &\rightarrow U_{02} = \{(x, y): 0 < x^2 + y^2 < 1, -\pi - \delta < \theta(x, y) < \delta\}, \\ \varphi_{21}(r, \theta) &= \begin{cases} (r, \theta): & (0, 1] \times (-\delta, \delta) \rightarrow (0, 1] \times (-\delta, \delta), \\ (r, \theta - 2\pi): & (0, 1] \times (\pi - \delta, \pi + \delta) \rightarrow (0, 1] \times (-\pi - \delta, -\pi + \delta). \end{cases}\end{aligned}$$

These are continuously differentiable. Therefore the disk is a differentiable manifold, with the unit circle as the boundary

$$\partial M = \sigma_1(0 \times (-\delta, \pi + \delta)) \cup \sigma_2((-\pi - \delta, \delta)) = \{(x, y): x^2 + y^2 = 1\}.$$

The idea of the example can be extended to show that the ball in  $\mathbb{R}^n$  is a manifold with the sphere  $S^{n-1}$  as the boundary. See Exercise 14.16.

**Example 14.1.9 (Möbius band).** Consider the circle  $\gamma(\theta) = (\cos \theta, \sin \theta, 0)$  lying in the  $(x, y)$ -plane in  $\mathbb{R}^3$ . At each point  $\gamma(\theta)$  of the circle, let  $I_\theta$  be the interval that is perpendicular to the circle, has angle  $\frac{1}{2}\theta$  from the  $(x, y)$ -plane, has length 1 and has the point  $\gamma(\theta)$  as the center. Basically  $I_\theta$  moves along the circle while rotating at half speed. The union  $M = \cup I_\theta$  is the Möbius band.

To find an atlas for the Möbius band, we note that the direction of  $I_\theta$  is

$$v(\theta) = \left( \cos \frac{1}{2}\theta \cos \theta, \cos \frac{1}{2}\theta \sin \theta, \sin \frac{1}{2}\theta \right).$$

Then we use Exercise 14.3 for the circular direction and get charts

$$\begin{aligned} \sigma_1(\theta, t) &= \gamma(\theta) + tv(\theta), & (\theta, t) &\in U_1 = (-\delta, \pi + \delta) \times [-0.5, 0.5], \\ \sigma_2(\theta, t) &= \gamma(\theta) + tv(\theta), & (\theta, t) &\in U_2 = (-\pi - \delta, \delta) \times [-0.5, 0.5]. \end{aligned}$$

Here we further relaxed the choice of  $U_i$  compared with Example 14.1.8, again to make our formulae more manageable. Strictly speaking, we should split  $\sigma_1$  into two charts

$$\sigma_{1+}(\theta, t) = \sigma_1(\theta, t - 0.5), \quad \sigma_{1-}(\theta, t) = \sigma_1(\theta, 0.5 - t), \quad (\theta, t) \in (-\delta, \pi + \delta) \times [0, 1],$$

and do the same to  $\sigma_2$ .

The transition has two parts corresponding to  $(-\delta, \delta)$  and  $(\pi - \delta, \pi + \delta)$

$$\varphi_{21}(\theta, t) = \begin{cases} (\theta, t): & (-\delta, \delta) \times [-0.5, 0.5] \rightarrow (-\delta, \delta) \times [-0.5, 0.5], \\ (\theta - 2\pi, -t): & (\pi - \delta, \pi + \delta) \times [-0.5, 0.5] \rightarrow (-\pi - \delta, -\pi + \delta) \times [-0.5, 0.5]. \end{cases}$$

The transition map is continuously differentiable, making the Möbius band into a manifold with boundary

$$\gamma(\theta) + \frac{1}{2}v(\theta) = \left( \left(1 + \frac{1}{2} \cos \frac{1}{2}\theta\right) \cos \theta, \left(1 + \frac{1}{2} \cos \frac{1}{2}\theta\right) \cos \frac{1}{2}\theta \sin \theta, \frac{1}{2} \sin \frac{1}{2}\theta \right).$$

The whole boundary is a circle obtained by moving  $\theta$  by  $4\pi$ , or twice around the circle  $\gamma$ .

**Exercise 14.15.** Let  $f$  be a continuously differentiable function on an open subset  $U$  of  $\mathbb{R}^n$ . Show that  $M = \{(\vec{x}, y) : y \leq f(\vec{x}), \vec{x} \in U\}$  is a manifold with boundary.

**Exercise 14.16.** Show that the unit ball  $B^n$  is a manifold with the unit sphere  $S^{n-1}$  as the boundary.

**Exercise 14.17.** Each point  $\theta$  of the real projective space  $\mathbb{R}P^n$  is a 1-dimensional subspace of  $\mathbb{R}^{n+1}$ . Let  $I_\theta$  be the interval of vectors in  $\theta$  of length  $\leq 1$ . Show that

$$M = \{(\theta, v) : \theta \in \mathbb{R}^{n+1}, v \in I_\theta\}$$

is a manifold with boundary. Moreover, show that the boundary is the sphere  $S^n$ .

**Exercise 14.18.** Repeat Exercise 14.17 for the complex projective space.

**Exercise 14.19.** Suppose  $M$  is a manifold with boundary. Prove that  $M - \partial M$  is a manifold without boundary.

**Exercise 14.20.** Suppose  $M$  and  $N$  are manifolds with boundaries. Prove that the product  $M \times N$  (see Exercise 14.13) is also a manifold, with boundary  $\partial M \times N \cup M \times \partial N$ .

## 14.2 Topology of Manifold

The rigorous definition of manifold requires further topological conditions. While a topology is usually defined as the collection of (open) subsets satisfying three axioms, it is more natural for us to start with the concept of limit. In fact, the whole theory of point set topology can be developed based on four axioms for limits<sup>35</sup>.

### Sequence Limit

The limit of a sequence in a manifold can be defined by using a chart to pull the sequence to Euclidean space. We notice that, in general, the chart can only be applied to later terms in converging sequence.

**Definition 14.2.1.** A sequence of points  $x_k$  in a manifold  $M$  *converges* to  $x$ , if there is a chart  $\sigma: U \rightarrow M$ , such that

- $x \in \sigma(U)$ ,
- $x_k \in \sigma(U)$  for sufficiently large  $k$ ,
- $\sigma^{-1}(x_k)$  converges to  $\sigma^{-1}(x)$  in  $U$ .

The first two items in the definition allows us to pull the sequence and the limit to the Euclidean space. Then we know the meaning of  $\lim \sigma^{-1}(x_k) = \sigma^{-1}(x)$  in  $U \subset \mathbb{R}^n$ . The following implies that the definition of limit is (in certain sense) independent of the choice of the chart.

**Lemma 14.2.2.** Suppose  $\lim x_k = x$  according to a chart  $\sigma: U \rightarrow M$ . If  $\tau: V \rightarrow M$  is another chart satisfying  $x \in \tau(V)$ , then  $\lim x_k = x$  according to the chart  $\tau$ .

*Proof.* Let  $M_\sigma = \sigma(U)$  and  $M_\tau = \tau(V)$ . Then by the definition of manifold,  $U_{\sigma\tau} = \sigma^{-1}(M_\sigma \cap M_\tau)$  and  $U_{\tau\sigma} = \tau^{-1}(M_\sigma \cap M_\tau)$  are open subsets of Euclidean space, and the transition map  $\varphi = \tau^{-1} \circ \sigma: U_{\sigma\tau} \rightarrow U_{\tau\sigma}$  is continuous.

By the assumption, we have  $x \in M_\sigma \cap M_\tau$ . This implies  $\sigma^{-1}(x) \in U_{\sigma\tau}$ . Since  $U_{\sigma\tau}$  is open and  $\lim \sigma^{-1}(x_k) = \sigma^{-1}(x) \in U_{\sigma\tau}$ , by Proposition 6.4.2, we know  $\sigma^{-1}(x_k) \in U_{\sigma\tau}$  for sufficiently large  $k$ . Then  $\tau^{-1}(x_k) = \varphi(\sigma^{-1}(x_k)) \in U_{\tau\sigma} \subset V$  for sufficiently large  $k$ . This implies the second item in the definition according to  $\tau$ . Moreover, by what we know about the continuity of the map  $\varphi$  between Euclidean spaces, we have

$$\lim \tau^{-1}(x_k) = \varphi(\lim \sigma^{-1}(x_k)) = \varphi(\sigma^{-1}(x)) = \tau^{-1}(x).$$

This verifies the third item in the definition according to  $\tau$ . □

<sup>35</sup>A good reference is Chapter 2 of *General Topology* by John Kelley.

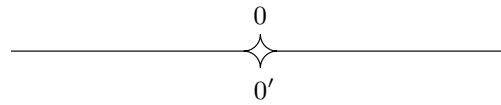
**Example 14.2.1** (Line with two origins). The *line with two origins*

$$M = (-\infty, 0) \sqcup \{0, 0'\} \sqcup (0, +\infty)$$

is covered by two charts

$$\begin{aligned}\sigma: \mathbb{R} &\rightarrow \mathbb{R} = (-\infty, 0) \sqcup \{0\} \sqcup (0, +\infty), \\ \sigma': \mathbb{R} &\rightarrow \mathbb{R}' = (-\infty, 0) \sqcup \{0'\} \sqcup (0, +\infty).\end{aligned}$$

The transition map is the identity map on  $\mathbb{R} - \{0\} = (-\infty, 0) \sqcup (0, +\infty)$ .



**Figure 14.2.1.** *Line with two origins.*

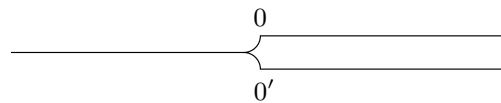
By Definition 14.2.1, we have  $\lim \frac{1}{k} = 0$  (according to  $\sigma$ ) and  $\lim \frac{1}{k} = 0'$  (according to  $\sigma'$ ). This shows that a sequence may converge to two different limits. This is not a counterexample to Lemma 14.2.2 because the lemma claims the well-defined limit only for the charts containing the limit point  $x$ . In the line with two origins, the chart  $\sigma'$  does not contain the first limit point 0.

**Exercise 14.21.** Prove that in a product manifold (see Exercise 14.13), a sequence  $(x_k, y_k)$  converges to  $(x, y)$  if and only if  $x_k$  converges to  $x$  and  $y_k$  converges to  $y$ .

**Exercise 14.22.** Show that 0 and  $0'$  are all the limits of the sequence  $\frac{1}{k}$  in the line with two origins.

**Exercise 14.23.** Construct the line with infinitely many origins, show that it is a manifold, and show that a sequence may have infinitely many limits.

**Exercise 14.24** (Forked line). Let  $[0, +\infty)'$  be a separate copy of  $[0, +\infty)$ . For  $x \geq 0$ , we denote by  $x$  the number in  $[0, +\infty)$ , and by  $x'$  the number in the copy  $[0, +\infty)'$ . Show that the *forked line*  $M = (-\infty, 0) \sqcup [0, +\infty) \sqcup [0, +\infty)'$  is a manifold and study the limit of sequences in the forked line.



**Figure 14.2.2.** *Forked line.*

**Exercise 14.25.** Construct a surface (i.e., 2-dimensional manifold) in which a sequence may not have unique limit.

## Compact, Closed, and Open Subsets

We use sequence limit to define the usual topological concepts, modelled on Definitions 6.3.2 and 6.3.5, and Proposition 6.4.2.

**Definition 14.2.3.** A subset of a manifold is *open* if any sequence converging to a point in the subset lies in the subset for sufficiently large index.

**Definition 14.2.4.** A subset of a manifold is *closed* if the limit of any convergent sequence in the subset still lies in the subset.

**Definition 14.2.5.** A subset of a manifold is *compact* if any sequence in the subset has a convergent subsequence, and the limit of the subsequence still lies in the subset.

In Sections 6.3 and 6.4, we saw that some topological properties can be derived by formal argument on sequence limit. In fact, the only property of the sequence limit we used is that, if a sequence converges, then any subsequence converges to the same limit. The following are these properties, and the formal argument based on sequence limit is repeated.

**Proposition 14.2.6.** *Open subsets have the following properties.*

1.  $\emptyset$  and  $M$  are open.
2. Unions of open subsets are open.
3. Finite intersections of open subsets are open.

*Proof.* The first property is trivially true.

Suppose  $U_i$  are open, and a sequence  $x_k$  converges to  $x \in \cup U_i$ . Then  $x$  is in some  $U_i$ . By  $U_i$  open, we have  $x_k \in U_i$  for sufficiently large  $k$ . Then by  $U_i \subset \cup U_i$ , we have  $x_k \in U$  for sufficiently large  $k$ .

Suppose  $U$  and  $V$  are open, and a sequence  $x_k$  converges to  $x \in U \cap V$ . By  $x \in U$  and  $U$  open, there is  $K_1$ , such that  $x_k \in U$  for  $k > K_1$ . By  $x \in V$  and  $V$  open, there is  $K_2$ , such that  $x_k \in V$  for  $k > K_2$ . Then  $x_k \in U \cap V$  for  $k > \max\{K_1, K_2\}$ .  $\square$

**Proposition 14.2.7.** *Closed subsets have the following properties.*

1.  $\emptyset$  and  $M$  are closed.
2. Intersections of closed subsets are closed.
3. Finite unions of closed subsets are closed.

*Proof.* The first property is trivially true.

Suppose  $C_i$  are closed, and a sequence  $x_k \in \cap C_i$  converges to  $x$ . Then the sequence  $x_k$  lies in each  $C_i$ . Then by  $C_i$  closed, the limit  $x \in C_i$ . Therefore  $x \in \cap C_i$ .

Suppose  $C$  and  $D$  are closed, and a sequence  $x_k \in C \cup D$  converges to  $x$ . Then there must be infinitely many  $x_k$  in either  $C$  or  $D$ . In other words, there is a subsequence in either  $C$  or  $D$ . If  $C$  contains a subsequence  $x_{k_p}$ , then by  $x = \lim x_{k_p}$  and  $C$  closed, we get  $x \in C$ . If  $D$  contains a subsequence, then we similarly get  $x \in D$ . Therefore  $x \in C \cup D$ .  $\square$

**Proposition 14.2.8.** *A subset is open if and only if its complement is closed.*

*Proof.* Suppose  $U$  is an open subset of  $M$ , and a sequence  $x_k \in M - U$  converges to  $x$ . If  $x \notin M - U$ , then  $x \in U$ . By  $x_k$  converging to  $x$  and  $U$  open, we have  $x_k \in U$  for sufficiently large  $k$ . This contradicts to all  $x_k \in M - U$ . Therefore we must have  $x \in M - U$ .

Suppose  $C$  is a closed subset of  $M$ , and a sequence  $x_k$  converges to  $x \in M - C$ . If it is not the case that  $x_k \in M - C$  for sufficiently large  $k$ , then there is a subsequence  $x_{k_p} \notin M - C$ . This means  $x_{k_p} \in C$ . Since  $C$  is closed, the limit of the subsequence, which is also the limit  $x$  of the sequence, must also lie in  $C$ . Since this contradicts to the assumption  $x \in M - C$ , we conclude that  $x_k \in M - C$  for sufficiently large  $k$ .  $\square$

**Exercise 14.26.** Prove that the boundary of a manifold is a closed subset.

**Exercise 14.27.** Suppose  $U_i$  are open subsets of  $M$  and  $M = \cup U_i$ . Prove that  $U \subset M$  is open if and only if all  $U \cap U_i$  are open. Can you find similar statement for closed subsets?

**Exercise 14.28.** Prove that the union of two compact subsets is compact.

**Exercise 14.29.** Prove that any closed subset of a compact set is compact.

**Exercise 14.30.** In the product of two manifolds (see Exercise 14.13), prove that the product of two open (closed, compact) subsets is still open (closed, compact).

**Exercise 14.31.** Describe open, closed and compact subsets of the line with two origins in Example 14.2.1.

The following characterises open subsets in terms of charts.

**Proposition 14.2.9.** *Let  $\{\sigma_i: U_i \rightarrow M_i\}$  be an atlas of a manifold  $M$ . Then all  $M_i$  are open, and a subset  $B \subset M$  is open if and only if  $\sigma_i^{-1}(B \cap M_i)$  is open in  $\mathbb{R}^n$  for all  $i$ .*

*Proof.* The openness of  $M_i$  means that, if a sequence  $x_k$  converges to  $x \in M_i$ , then  $x_k \in M_i$  for sufficiently large  $k$ . Although the definition of  $x_k$  converging to  $x$  requires the choice of a chart, Lemma 14.2.2 shows that we can use any chart

containing  $x$ . In particular, we may apply Lemma 14.2.2 to the chart  $\sigma_i$  (in place of  $\tau$  in the lemma). The conclusion is that  $x_k \in M_i$  for sufficiently large  $k$ , and  $\sigma_i^{-1}(x_k)$  converges to  $\sigma_i^{-1}(x)$ . The first part of the conclusion means the openness of  $M_i$ .

Once we know  $M_i$  are open, it then follows from the general property in Proposition 14.2.6 that  $B$  is open if and only if  $B \cap M_i$  is open for all  $i$ . See Exercise 14.27. For  $A = B \cap M_i \subset M_i$ , it remains to show that  $A$  is open in  $M$  if and only if  $\sigma_i^{-1}(A)$  is open in  $\mathbb{R}^n$ .

Suppose  $\sigma_i^{-1}(A)$  is open and  $\lim x_k = x \in A \subset M_i$ . Then  $\sigma_i$  contains  $x$  and we may apply Lemma 14.2.2 to get  $x_k \in M_i$  for sufficiently large  $k$ , and  $\lim \sigma_i^{-1}(x_k) = \sigma_i^{-1}(x)$  in  $\mathbb{R}^n$ . Since  $\sigma_i^{-1}(A)$  is open, the limit implies  $\sigma_i^{-1}(x_k) \in \sigma_i^{-1}(A)$  for sufficiently large  $k$ . Then we conclude that  $x_k \in A$  for sufficiently large  $k$ . This proves the openness of  $A$ .

Conversely, suppose  $A$  is open and  $\lim \vec{u}_k = \vec{u} \in \sigma_i^{-1}(A)$ . Since  $\vec{u} \in \sigma_i^{-1}(A) \subset U_i$  and  $U_i$  is open in  $\mathbb{R}^n$ , we get  $\vec{u}_k \in U_i$  for sufficiently large  $k$ . Up to deleting finitely many terms, therefore, we may assume all  $\vec{u}_k \in U_i$ , so that  $\sigma_i$  can be applied to  $\vec{u}_k$  and  $\vec{u}$ . Then we may verify  $\lim \sigma_i(\vec{u}_k) = \sigma_i(\vec{u})$  by applying the definition to the chart  $\sigma_i$ . The assumption  $\lim \vec{u}_k = \vec{u}$  becomes the third item in the definition, and the limit is verified. Now the assumption  $\vec{u} \in \sigma_i^{-1}(A)$  means  $\sigma_i(\vec{u}) \in A$ . Then by the openness of  $A$ , the limit implies  $\sigma_i(\vec{u}_k) \in A$  for sufficiently large  $k$ . In other words, we get  $\vec{u}_k \in \sigma_i^{-1}(A)$  for sufficiently large  $k$ . This proves the openness of  $\sigma_i^{-1}(A)$ .  $\square$

**Exercise 14.32.** Prove that an open subset of a manifold is also a manifold.

## Rigorous Definition of Manifold

With the additional knowledge about the topology of manifolds, we are able to amend Definition 14.1.1 and give the rigorous definition.

**Definition 14.2.10.** A *differentiable manifold* of dimension  $n$  is a set  $M$  and an atlas with continuously differentiable transition maps, such that the following topological properties are satisfied.

1. *Hausdorff*: If  $x \neq y$  are distinct points in  $M$ , then there are disjoint open subsets  $B_x$  and  $B_y$ , such that  $x \in B_x$  and  $y \in B_y$ .
2. *Paracompact*: Every open cover has a locally finite refinement.

The reason for the Hausdorff property is the following result. Without the Hausdorff property, some pathological phenomenon may occur. In fact, Stokes' theorem will not hold (at least in its usual form) without the Hausdorff property. See Exercise 14.37.

**Proposition 14.2.11.** *In a manifold with Hausdorff property, the limit of a sequence is unique.*



*Proof.* Suppose  $\lim x_k = x$  and  $\lim x_k = y$ . If  $x \neq y$ , then we have open subsets  $B_x$  and  $B_y$  in the definition of Hausdorff property. By  $x \in B_x$ ,  $y \in B_y$  and the definition of open subsets, we have  $x_k \in B_x$  and  $x_k \in B_y$  for sufficiently large  $k$ . However, this is impossible because  $B_x$  and  $B_y$  are disjoint.  $\square$

The meaning of the paracompact property is less clear at this moment. In fact, we will not explain the full detail of the definition in this course. The concept will only be used when we define the integration on manifold, and the significance of the concept will be elaborated then.

**Exercise 14.33.** Prove that  $\mathbb{R}^n$  with its usual open subsets is Hausdorff.

**Exercise 14.34.** Prove that a manifold  $M$  with atlas is Hausdorff if and only if the following holds: If  $x \in M_i - M_j$  and  $y \in M_j - M_i$ , then there are open subsets  $U_x \subset U_i$ ,  $U_y \subset U_j$ , such that

$$x \in \sigma_i(U_x), \quad y \in \sigma_j(U_y), \quad \sigma_i(U_x) \cap \sigma_j(U_y) = \emptyset.$$

**Exercise 14.35.** Show that the line with two origins in Example 14.2.1 and the forked line in Exercise 14.24 are not Hausdorff.

**Exercise 14.36.** Suppose  $M$  and  $N$  are Hausdorff manifolds. Prove that the product manifold  $M \times N$  (see Exercise 14.13) is also a Hausdorff manifold.

**Exercise 14.37.** The “interval”  $I = [-1, 0) \sqcup [0, 1] \sqcup [0, 1]'$  in the forked line in Exercise 14.24 has left end  $-1$  and right ends  $1, 1'$ . Verify that the “fundamental theorem of calculus”

$$\int_I f' dx = \int_{-1}^0 f' dx + \int_0^1 f' dx + \int_{0'}^{1'} f' dx = f(1) + f(1') - f(-1)$$

holds for  $f(x) = x$  and does not hold for  $f(x) = x + 1$ .

## Continuous Map

The continuity of maps between manifolds cannot be defined using the usual  $\epsilon$ - $\delta$  language because there is no norm (or distance) on manifolds. We may define the continuity in terms of sequence limit.

**Definition 14.2.12.** A map  $F: M \rightarrow N$  between manifolds is *continuous* if  $\lim x_k = x$  implies  $\lim F(x_k) = F(x)$ .

**Example 14.2.2.** Consider a chart map  $\sigma: U \rightarrow M$ . The continuity of  $\sigma$  means that  $\lim \vec{u}_k = \vec{u}$  in  $U \subset \mathbb{R}^n$  implies  $\lim \sigma(\vec{u}_k) = \sigma(\vec{u})$ . By applying the definition of  $\lim \sigma(\vec{u}_k) = \sigma(\vec{u})$  to the chart  $\sigma$ , we see the first two items are clearly satisfied, and the third item is exactly  $\lim \vec{u}_k = \vec{u}$ . Therefore  $\sigma$  is continuous.

**Exercise 14.38.** Prove that the composition of continuous maps is continuous.

**Exercise 14.39.** For a chart  $\sigma: U \rightarrow M$ , prove that  $\sigma^{-1}: \sigma(U) \rightarrow U$  is also continuous.

The following characterises continuous maps in terms of charts.

**Proposition 14.2.13.** *For a map  $F: M \rightarrow N$  between manifolds, the following are equivalent.*

- $F$  is continuous.
- If  $B \subset N$  is open, then  $F^{-1}(B)$  is open.
- For any  $x \in M$ , there are charts  $\sigma: U \rightarrow M_\sigma \subset M$ ,  $\tau: V \rightarrow N_\tau \subset N$ , such that  $x \in M_\sigma$ ,  $F(M_\sigma) \subset N_\tau$ , and the composition  $\tau^{-1} \circ F \circ \sigma: U \rightarrow V$  is continuous.

Note that the composition  $\tau^{-1} \circ F \circ \sigma$  is defined precisely when  $F(M_\sigma) \subset N_\tau$ , and is a map between Euclidean spaces. The proof shows that we may start with any two charts  $\sigma$  and  $\tau$  around  $x$  and  $F(x)$ , and then restrict to a smaller part of  $\sigma$  that still contains  $x$ , such that the restriction  $\sigma|$  satisfies  $F(M_{\sigma|}) \subset N_\tau$ .

*Proof.* Suppose  $F: M \rightarrow N$  is continuous,  $B \subset N$  is open, and  $\lim x_k = x \in F^{-1}(B)$ . We wish to show that  $x_k \in F^{-1}(B)$  for sufficiently large  $k$ . Note that the continuity of  $F$  implies  $\lim F(x_k) = F(x)$ . Moreover,  $x \in F^{-1}(B)$  means  $F(x) \in B$ . Then the openness of  $B$  implies that  $F(x_k) \in B$  for sufficiently large  $k$ . This means that  $x_k \in F^{-1}(B)$  for sufficiently large  $k$ . This completes the proof that the first statement implies the second.

Next, suppose  $F^{-1}(B)$  is open for any open  $B$ . Let  $\sigma: U \rightarrow M_\sigma \subset M$  and  $\tau: V \rightarrow N_\tau \subset M$  be charts around  $x$  and  $F(x)$ . By Proposition 14.2.9,  $M_\sigma$  and  $N_\tau$  are open subsets. Then the preimage  $F^{-1}(N_\tau)$  is open, and the intersection  $M_\sigma \cap F^{-1}(N_\tau)$  is also open. By Example 14.2.2, we know  $\sigma$  is continuous. By what we just proved, the preimage  $U' = \sigma^{-1}(M_\sigma \cap F^{-1}(N_\tau))$  is an open subset of  $U$ . Then the restriction  $\sigma|_{U'}: U' \rightarrow M'_\sigma \subset M$  is still a (smaller) chart around  $x$ , such that  $M'_\sigma = \sigma(U') = M_\sigma \cap F^{-1}(N_\tau) \subset F^{-1}(N_\tau)$ . This is the same as  $F(M'_\sigma) \subset N_\tau$ .

So by replacing  $\sigma$  with  $\sigma|_{U'}$ , we may assume that  $F(M_\sigma) \subset N_\tau$  is satisfied. By Example 14.2.2 and Exercises 14.38 and 14.39, the composition  $\tau^{-1} \circ F \circ \sigma$  is continuous. This completes the proof that the second statement implies the third.

Finally, we assume that the third statement holds. For  $\lim x_k = x$ , we wish to show that  $\lim F(x_k) = F(x)$ . Let  $\sigma, \tau$  be the charts around  $x$  and  $F(x)$  in the third statement. Then the composition  $\tau^{-1} \circ F \circ \sigma: U \rightarrow V$  is continuous. By Example 14.2.2 and Exercise 14.39, we know  $\sigma^{-1}: M_\sigma \rightarrow U$  and  $\tau: V \rightarrow N$  are continuous. Then by Exercise 14.38, the composition  $F|_{M_\sigma} = \tau \circ (\tau^{-1} \circ F \circ \sigma) \circ \sigma^{-1}: M_\sigma \rightarrow N$  is also continuous. By  $\lim x_k = x \in M_\sigma$  and the openness of  $M_\sigma$ , we have  $x_k \in M_\sigma$  for sufficiently large  $k$ . Therefore  $\lim x_k = x$  happens inside  $M_\sigma$ . Applying the continuity of  $F|_{M_\sigma}: M_\sigma \rightarrow N$  to the limit, we conclude that  $\lim F(x_k) = F(x)$ . This completes the proof that the third statement implies the first.  $\square$

**Example 14.2.3.** A continuous curve in a manifold  $M$  is a continuous map  $\gamma: I \rightarrow M$  from an interval  $I$ . The continuity means that if  $\gamma(t) \in \sigma(U)$  for a chart  $\sigma$  of  $M$ , then for sufficiently small  $\delta$ , we have  $\gamma(t - \delta, t + \delta) \subset \sigma(U)$ , and  $\sigma^{-1} \circ \gamma: (t - \delta, t + \delta) \rightarrow U$  is continuous.

**Example 14.2.4.** A continuous function on a manifold  $M$  is a continuous map  $f: M \rightarrow \mathbb{R}$ . By taking the identity map  $\mathbb{R} \rightarrow \mathbb{R}$  as the chart  $\tau$  in Proposition 14.2.13, we find that the continuity means that, for any chart  $\sigma: U \rightarrow M$ , the composition  $f \circ \sigma$  is a continuous function on  $U$ .

**Example 14.2.5.** The canonical map

$$F(\vec{x}) = [\vec{x}]: S^n \rightarrow \mathbb{R}P^n, \quad \vec{x} = (x_0, x_1, \dots, x_n)$$

sends a unit length vector to the 1-dimensional subspace spanned by the vector. For  $x_0 > 0$ , we have a chart for  $S^n$

$$\sigma_0(\vec{u}) = \left( \sqrt{1 - \|\vec{u}\|_2^2}, \vec{u} \right): U = \{\vec{u} \in \mathbb{R}^n: \|\vec{u}\|_2 < 1\} \rightarrow S^n.$$

For  $x_0 \neq 0$ , we have a chart for  $\mathbb{R}P^n$

$$\tau_0(\vec{w}) = [1, \vec{w}]: V = \mathbb{R}^n \rightarrow \mathbb{R}P^n.$$

The composition

$$\tau_0^{-1} \circ F \circ \sigma_0(\vec{u}) = \frac{\vec{u}}{\sqrt{1 - \|\vec{u}\|_2^2}}: U \rightarrow V$$

is defined on the whole  $U$  and is continuous.

Similar charts  $\sigma_i$  and  $\tau_i$  can be found for  $x_i > 0$  and for  $x_i \neq 0$ , as well as for  $x_i < 0$  and for  $x_i \neq 0$ . In all cases, the composition  $\tau_i^{-1} \circ F \circ \sigma_i$  is continuous. We conclude that  $F$  is continuous.

**Exercise 14.40.** Construct a canonical map  $S^{2n+1} \rightarrow \mathbb{C}P^n$  similar to Example 14.2.5 and show that the map is continuous.

**Exercise 14.41.** Prove that a map  $F: M \rightarrow N$  is continuous if and only if for any closed subset  $C$  of  $N$ ,  $F^{-1}(C)$  is closed subset of  $M$ .

**Exercise 14.42.** Suppose  $\{\sigma_i\}$  is an atlas of  $M$ . Prove that a map  $F: M \rightarrow N$  is continuous if and only if the compositions  $F \circ \sigma_i: U_i \rightarrow N$  are continuous for each  $i$ .

**Exercise 14.43.** Suppose  $F: M \rightarrow N$  is continuous. Prove that there are atlases  $\{\sigma_i: U_i \rightarrow M_i \subset M\}$  and  $\{\tau_i: V_i \rightarrow N_i \subset N\}$  with the same index set, such that  $F(M_i) \subset N_i$ , and the composition  $\tau_i^{-1} \circ F \circ \sigma_i: U_i \rightarrow V_i$  is continuous for each  $i$ . Can you formulate the converse statement?

## 14.3 Tangent and Cotangent

The tangent space of a submanifold  $M \subset \mathbb{R}^N$  at a point  $\vec{x}_0 \in M$  is defined in Section 8.4 to be the collection of tangent vectors of all the differentiable curves  $\gamma$  in  $M$  passing through  $\vec{x}_0$

$$T_{\vec{x}_0}M = \{\gamma'(0): \gamma(t) \in M \text{ for all } t, \gamma(0) = \vec{x}_0\}.$$

Here the curve  $\gamma$  is regarded as a map into  $\mathbb{R}^N$ , and  $\gamma'(0)$  is a vector in  $\mathbb{R}^N$ . By the chain rule, we also have

$$\left. \frac{d}{dt} \right|_{t=0} f(\gamma(t)) = f'(\vec{x}_0)(\gamma'(0)).$$

Since a manifold may not be inside  $\mathbb{R}^N$  (despite many ways of embedding a manifold in the Euclidean space), we cannot use the vectors  $\gamma'(0) \in \mathbb{R}^N$  in general. However, the left side still makes sense in general and can be considered as the “effect” of the tangent vector (i.e., the directional derivative) on the function  $f$ .

Example 14.2.3 describes continuous curves in a manifold. If  $M$  is a differentiable manifold and  $\sigma^{-1} \circ \gamma: (t - \delta, t + \delta) \rightarrow U$  is always differentiable, then we say that the curve is *differentiable*. See Example 14.2.3 and Exercise 14.65.

Example 14.2.4 describes continuous functions in a manifold. If  $M$  is a differentiable manifold and  $f \circ \sigma$  (defined on an open subset of Euclidean space) is always differentiable, then we say that the function is *differentiable*. See Example 14.2.4 and Exercise 14.64. Since we will only be interested at the linear approximations of a manifold near a point, we only need the function to be defined near the point, and  $f \circ \sigma$  is differentiable at the point.

Let  $M$  be a differentiable manifold, and  $x_0 \in M$ . For any differentiable curve  $\gamma$  at  $x_0$  (i.e., satisfying  $\gamma(0) = x_0$ ) and any differentiable function  $f$  at  $x_0$  (i.e., defined near  $x_0$ ), we introduce the pairing

$$\langle \gamma, f \rangle = \left. \frac{d}{dt} \right|_{t=0} f(\gamma(t)). \quad (14.3.1)$$

**Definition 14.3.1.** Let  $M$  be a differentiable manifold and  $x_0 \in M$ .

- Two differentiable curves  $\gamma_1$  and  $\gamma_2$  at  $x_0$  are equivalent if

$$\langle \gamma_1, f \rangle = \langle \gamma_2, f \rangle \text{ for all differentiable functions } f \text{ at } x_0.$$

The equivalence class of a curve  $\gamma$  is denoted  $[\gamma]$  and called a *tangent vector*. All the equivalence classes  $[\gamma]$  form the *tangent space*  $T_{x_0}M$  of  $M$  at  $x_0$ .

- Two differentiable functions  $f_1$  and  $f_2$  at  $x_0$  are equivalent if

$$\langle \gamma, f_1 \rangle = \langle \gamma, f_2 \rangle \text{ for all differentiable curves } \gamma \text{ at } x_0.$$

The equivalence class of a function  $f$  is denoted  $d_{x_0}f = [f]$  and called the *differential* of  $f$  at  $x_0$ . All the differentials  $d_{x_0}f$  form the *cotangent space*  $T_{x_0}^*M$  of  $M$  at  $x_0$ .

It is easy to see that the equivalences in the definition are indeed equivalence relations. Then the pairing (14.3.1) really means a pairing between equivalence classes,

$$\langle [\gamma], d_{x_0}f \rangle = \left. \frac{d}{dt} \right|_{t=0} f(\gamma(t)): T_{x_0}M \times T_{x_0}^*M \rightarrow \mathbb{R}. \quad (14.3.2)$$

For a tangent vector  $X = [\gamma] \in T_{x_0}M$  and a *cotangent vector*  $\omega = d_{x_0}f \in T_{x_0}^*M$ , we also denote the pairing by

$$\langle X, \omega \rangle = X(f) = \omega(X).$$

The elements are called vectors because we will show that both  $T_{x_0}M$  and  $T_{x_0}^*M$  are vector spaces, and  $\langle X, \omega \rangle$  is a dual pairing between the two vector spaces. In writing  $X(f)$ , we mean the derivative of  $f$  in the direction of  $X$ . In writing  $\omega(X)$ , we mean the linear functional  $\omega$  evaluated at vector  $X$ .

We denote the equivalence class  $[f]$  by  $d_{x_0}f$ . Similarly, we really should denote the equivalence class  $[\gamma]$  by the tangent vector  $\gamma'(0)$  of the curve. However, we cannot use the notation  $\gamma'(0)$  until the concept of the derivative of differentiable map between differentiable manifolds is introduced in Definition 14.4.2. Then the notation  $\gamma'(0)$  will be explained in Example 14.4.3.

## Tangent and Cotangent Vector Spaces

The vector space structure on the cotangent space  $T_{x_0}^*M$  can be easily defined by the linear combination of functions

$$a[f] + b[g] = [af + bg].$$

Using the differentials of functions, this can also be written as

$$d_{x_0}(af + bg) = ad_{x_0}f + bd_{x_0}g.$$

It follows easily from (14.3.1) that

$$\langle \gamma, af + bg \rangle = a\langle \gamma, f \rangle + b\langle \gamma, g \rangle.$$

Since the right side depends only on the equivalence classes of  $f$  and  $g$ , this implies that vector space structure on  $T_{x_0}^*M$  is well defined. This also implies that the pairing is linear in the cotangent vector

$$\langle X, a\omega + b\rho \rangle = a\langle X, \omega \rangle + b\langle X, \rho \rangle. \quad (14.3.3)$$

**Exercise 14.44.** Prove that  $[fg] = f(x_0)[g] + g(x_0)[f]$ , or  $d_{x_0}(fg) = f(x_0)d_{x_0}g + g(x_0)d_{x_0}f$ .

**Exercise 14.45.** Let  $\lambda(t)$  be a single variable function. Prove that  $[\lambda \circ f] = \lambda'(f(x_0))[f]$ , or  $d_{x_0}(\lambda \circ f) = \lambda'(f(x_0))d_{x_0}f$ .

The vector space structure for the tangent space  $T_{x_0}M$  is more complicated because it is not clear how to add two curves together (and still get a curve in  $M$ ). (The scalar multiplication can be achieved by changing  $\gamma(t)$  to  $\gamma(ct)$ .) We need to use charts to define the vector space structure on  $T_{x_0}M$ .

Let  $\sigma: U \rightarrow M$  be a chart, and  $x_0 = \sigma(\vec{u}_0)$ . The chart translates a curve  $\gamma$  in  $M$  to a curve  $\phi = \sigma^{-1} \circ \gamma$  in  $U$  and translates a function  $f$  on  $M$  to a function  $g = f \circ \sigma$  on  $U$ . Then  $f(\gamma(t)) = g(\phi(t))$  and the pairing (14.3.1) is translated into

$$\langle \gamma, f \rangle = \left. \frac{d}{dt} \right|_{t=0} g(\phi(t)) = g'(\vec{u}_0)(\phi'(0)). \quad (14.3.4)$$

Here the second equality is the classical chain rule.

The translation suggests the following maps (the notations will be explained in Section 14.4)

$$\begin{aligned} (\sigma^{-1})'(x_0): T_{x_0}M &\rightarrow \mathbb{R}^n, & [\gamma] &\mapsto \phi'(0) = (\sigma^{-1} \circ \gamma)'(0); \\ \sigma'(\vec{u}_0)^*: T_{x_0}^*M &\rightarrow (\mathbb{R}^n)^*, & [f] &\mapsto g'(0) = (f \circ \sigma)'(0). \end{aligned}$$

We will show that the maps are well-defined and are one-to-one correspondences. We will use the first map to define the vector space structure of  $T_{x_0}M$ . We will show that the second map is an isomorphism of vector spaces.

To show the maps are one-to-one correspondences, we find their inverse maps. For a vector  $\vec{v} \in \mathbb{R}^n$ , the most natural curve at  $\vec{u}_0$  in the direction of  $\vec{v}$  is the straight line  $\phi(t) = \vec{u}_0 + t\vec{v}$ . The corresponding curve in  $M$  is  $\gamma(t) = \sigma(\vec{u}_0 + t\vec{v})$ . This suggests the possible inverse to  $(\sigma^{-1})'(x_0)$

$$\sigma'(\vec{u}_0): \mathbb{R}^n \rightarrow T_{x_0}M, \quad \vec{v} \mapsto [\sigma(\vec{u}_0 + t\vec{v})].$$

On the other hand, a linear functional  $l \in (\mathbb{R}^n)^*$  is a function on  $\mathbb{R}^n$ . The corresponding function on  $M$  is  $f = l \circ \sigma^{-1}$ . This suggests the possible inverse to  $\sigma'(\vec{u}_0)^*$

$$(\sigma^{-1})'(x_0)^*: (\mathbb{R}^n)^* \rightarrow T_{x_0}^*M, \quad l \mapsto [l \circ \sigma^{-1}].$$

Again the notations for the two maps will be explained in Section 14.4.

**Proposition 14.3.2.** *Let  $\sigma: U \rightarrow M$  be a chart, and  $x_0 = \sigma(\vec{u}_0)$ . Then  $(\sigma^{-1})'(x_0)$  and  $\sigma'(\vec{u}_0)^*$  are well defined, and are invertible with respective inverse maps  $\sigma'(\vec{u}_0)$  and  $(\sigma^{-1})'(x_0)^*$ . Moreover,  $\sigma'(\vec{u}_0)^*$  and  $(\sigma^{-1})'(x_0)^*$  give an isomorphism of vector spaces.*

*Proof.* The equality (14.3.4) implies that  $\langle \gamma_1, f \rangle = \langle \gamma_2, f \rangle$  for all  $f$  is the same as  $g'(\vec{u}_0)(\phi'_1(0)) = g'(\vec{u}_0)(\phi'_2(0))$  for all  $g$ . By choosing all possible  $g$ ,  $g'(\vec{u}_0)$  can be any linear functional on  $\mathbb{R}^n$ . So the values of any linear functional at  $\phi'_1(0)$  and  $\phi'_2(0)$  are the same. This implies that  $\phi'_1(0) = \phi'_2(0)$  and proves that  $(\sigma^{-1})'(x_0)$  is well defined. Moreover, we note that  $\phi'_1(0) = \phi'_2(0)$  implies  $\langle \gamma_1, f \rangle = \langle \gamma_2, f \rangle$  for all  $f$ , so that  $[\gamma_1] = [\gamma_2]$ . This means that the map  $(\sigma^{-1})'(x_0)$  is injective.

The following shows that the composition  $(\sigma^{-1})'(x_0) \circ \sigma'(\vec{u}_0)$  is the identity

$$\begin{aligned} \vec{v} \in \mathbb{R}^n &\mapsto [\sigma(\vec{u}_0 + t\vec{v})] \in T_{x_0}M \\ &\mapsto (\sigma^{-1} \circ \sigma(\vec{u}_0 + t\vec{v}))'(0) = (\vec{u}_0 + t\vec{v})'(0) = \vec{v} \in \mathbb{R}^n. \end{aligned}$$

Combined with the injectivity of  $(\sigma^{-1})'(x_0)$ , we conclude that  $(\sigma^{-1})'(x_0)$  and  $\sigma'(\vec{u}_0)$  are inverse to each other.

By the similar argument, we can prove that  $\sigma'(\vec{u}_0)^*$  is well defined, and  $\sigma'(\vec{u}_0)^*$  and  $(\sigma^{-1})'(x_0)^*$  are inverse to each other. Moreover, the following shows that

$\sigma'(\vec{u}_0)^*$  is a linear transform

$$\begin{aligned}\sigma'(\vec{u}_0)^*(a_1[f_1] + a_2[f_2]) &= \sigma'(\vec{u}_0)^*([a_1f_1 + a_2f_2]) \\ &= ((a_1f_1 + a_2f_2) \circ \sigma)'(0) \\ &= (a_1(f_1 \circ \sigma) + a_2(f_2 \circ \sigma))'(0) \\ &= a_1(f_1 \circ \sigma)'(0) + a_2(f_2 \circ \sigma)'(0) \\ &= a_1\sigma'(\vec{u}_0)^*([f_1]) + a_2\sigma'(\vec{u}_0)^*([f_2]).\end{aligned}\quad \square$$

The pair of invertible maps  $(\sigma^{-1})'(x_0)$  and  $\sigma'(\vec{u}_0)$  can be used to translate the vector space structure on  $\mathbb{R}^n$  to a vector space structure on  $T_{x_0}M$ . This means that, for  $X, Y \in T_{x_0}M$ , we write

$$X = \sigma'(\vec{u}_0)(\vec{v}) = [\sigma(\vec{u}_0 + t\vec{v})], \quad Y = \sigma'(\vec{u}_0)(\vec{w}) = [\sigma(\vec{u}_0 + t\vec{w})], \quad \vec{v}, \vec{w} \in \mathbb{R}^n,$$

and then define

$$aX + bY = \sigma'(\vec{u}_0)(a\vec{v} + b\vec{w}) = [\sigma(\vec{u}_0 + t(a\vec{v} + b\vec{w}))].$$

**Proposition 14.3.3.** *The vector space structure on  $T_{x_0}M$  is well defined, and (14.3.2) is a dual pairing between  $T_{x_0}M$  and  $T_{x_0}^*M$ .*

*Proof.* By the vector space structure on  $T_{x_0}M$  being well defined, we mean the structure is independent of the choice of chart. Let  $\tau: V \rightarrow M$  be another chart, and  $x_0 = \tau(\vec{v}_0)$ . Then well defined means that the following composition should be linear

$$\mathbb{R}^n \xrightarrow{\sigma'(\vec{u}_0)} T_{x_0}M \xrightarrow{(\tau^{-1})'(x_0)} \mathbb{R}^n.$$

Let  $\varphi = \tau^{-1} \circ \sigma$  be the transition map between two charts. Then the composition

$$\begin{aligned}(\tau^{-1})'(x_0) \circ \sigma'(\vec{u}_0): \vec{v} \in \mathbb{R}^n &\mapsto [\sigma(\vec{u}_0 + t\vec{v})] \in T_{x_0}M \\ &\mapsto (\tau^{-1} \circ \sigma(\vec{u}_0 + t\vec{v}))'(0) = \varphi'(\vec{u}_0)(\vec{v}) \in \mathbb{R}^n\end{aligned}$$

is the derivative of the transition map and is therefore linear.

We already know that the pairing between  $T_{x_0}M$  and  $T_{x_0}^*M$  is linear in the cotangent vector by (14.3.3). The following shows that the pairing is also linear in the tangent vector

$$\begin{aligned}\langle aX + bY, [f] \rangle &= \left. \frac{d}{dt} \right|_{t=0} g(\vec{u}_0 + t(a\vec{v} + b\vec{w})) \\ &= g'(\vec{u}_0)(a\vec{v} + b\vec{w}) = ag'(\vec{u}_0)(\vec{v}) + bg'(\vec{u}_0)(\vec{w}) \\ &= a \left. \frac{d}{dt} \right|_{t=0} g(\vec{u}_0 + t\vec{v}) + b \left. \frac{d}{dt} \right|_{t=0} g(\vec{u}_0 + t\vec{w}) \\ &= a\langle X, [f] \rangle + b\langle Y, [f] \rangle.\end{aligned}$$

The definition of the equivalence classes  $[\gamma]$  and  $[f]$  implies that the pairing is non-singular (see Exercise 7.32). Therefore we get a dual pairing between  $T_{x_0}M$  and  $T_{x_0}^*M$ .  $\square$

**Example 14.3.1.** In Example 14.1.1, we explained that an open subset  $U \subset \mathbb{R}^n$  is a manifold, with an atlas given by the identity map chart  $U \rightarrow U$ . The isomorphism between the tangent space and  $\mathbb{R}^n$  induced by the chart is

$$\begin{aligned}\mathbb{R}^n &\rightarrow T_{\vec{u}_0}U, & \vec{v} &\mapsto [\vec{u}_0 + t\vec{v}], \\ T_{\vec{u}_0}U &\rightarrow \mathbb{R}^n, & [\gamma] &\mapsto \gamma'(0).\end{aligned}$$

This is consistent with our understanding of tangent vectors in Euclidean space. The isomorphism between the cotangent space and  $(\mathbb{R}^n)^*$  induced by the chart is

$$\begin{aligned}(\mathbb{R}^n)^* &\rightarrow T_{\vec{u}_0}^*U, & l &\mapsto [l], \\ T_{\vec{u}_0}^*U &\rightarrow (\mathbb{R}^n)^*, & [f] &\mapsto f'(\vec{u}_0).\end{aligned}$$

**Example 14.3.2.** For a submanifold  $M$  of  $\mathbb{R}^N$ , a chart  $\sigma: U \rightarrow M$  is a regular parameterization (see Section 8.4). We view  $\sigma: \mathbb{R}^n \rightarrow \mathbb{R}^N$  as a differentiable map between Euclidean spaces, and view a differentiable function on  $M$  as the restriction of a differentiable function  $f$  on  $\mathbb{R}^N$ . Then the isomorphism  $\sigma'(\vec{u}_0): \mathbb{R}^n \rightarrow T_{x_0}M$  sends  $\vec{v} \in \mathbb{R}^n$  to  $X = [\sigma(\vec{u}_0 + t\vec{v})] \in T_{\vec{x}_0}M$ , and we have

$$X(f) = \left. \frac{d}{dt} \right|_{t=0} (f(\sigma(\vec{u}_0 + t\vec{v}))) = f'(\vec{x}_0)(\sigma'(\vec{u}_0)(\vec{v})).$$

Here the first equality is the definition of  $X(f)$ , and the second equality is the usual chain rule for maps between Euclidean spaces (see Example 8.2.5). In particular,  $\sigma'(\vec{u}_0): \mathbb{R}^n \rightarrow \mathbb{R}^N$  on the right is not the isomorphism introduced before Proposition 14.3.2, and is the derivative linear transform introduced in Section 8.1. Therefore  $f'(\vec{x}_0)(\sigma'(\vec{u}_0)(\vec{v}))$  is the derivative of the multivariable function  $f$  (on  $\mathbb{R}^N$ ) in the direction of the vector  $\sigma'(\vec{u}_0)(\vec{v}) \in \mathbb{R}^N$ . The equality above then identifies  $X \in T_{\vec{x}_0}M$  with  $\sigma'(\vec{u}_0)(\vec{v}) \in \mathbb{R}^N$ , which leads to

$$T_{\vec{x}_0}M = \text{image}(\sigma'(\vec{u}_0): \mathbb{R}^n \rightarrow \mathbb{R}^N).$$

This recovers the tangent space in (8.4.1).

**Exercise 14.46.** Directly verify that the maps in Example 14.3.1 are inverse to each other.

**Exercise 14.47.** In Proposition 14.3.2, prove that  $\sigma'(\vec{u}_0)^*$  is well defined, and  $\sigma'(\vec{u}_0)^*$  and  $(\sigma^{-1})'(x_0)^*$  are inverse to each other.

**Exercise 14.48.** In Proposition 14.3.2, directly verify that  $(\sigma^{-1})'(x_0)^*$  is a linear transform.

**Exercise 14.49.** Suppose  $X_n \in T_{x_0}M$  is a sequence of tangent vectors and  $f_n$  is a sequence of functions at  $x_0$ . Use Exercise 7.24 to prove that, if  $X_n$  converges to  $X$  and  $df_n \in T_{x_0}^*M$  converges to  $df \in T_{x_0}^*M$ , then  $\lim X_n(f_n) = X(f)$ .

**Exercise 14.50.** Use Proposition 7.2.1 to prove the following.

1. A sequence of tangent vectors  $X_n \in T_{x_0}M$  converges to  $X \in T_{x_0}M$  if and only if  $\lim X_n(f) = X(f)$  for all differentiable functions  $f$ .
2. For a sequence of differentiable functions  $f_n$ , the corresponding sequence of cotangent vectors  $d_{x_0}f_n \in T_{x_0}^*M$  converges to  $d_{x_0}f \in T_{x_0}^*M$  if and only if  $\lim X(f_n) = X(f)$  for all tangent vectors  $X$ .



**Exercise 14.51.** Let  $X_t \in T_{x_0}M$  be a curve in the tangent space. Use Proposition 7.2.1 and Exercise 7.35 to prove that

$$\left. \frac{d}{dt} \right|_{t=0} X_t = Y \iff \left. \frac{d}{dt} \right|_{t=0} X_t(f) = Y(f) \text{ for all } f,$$

and

$$\int_a^b X_t dt = Y \iff \int_a^b X_t(f) dt = Y(f) \text{ for all } f.$$

**Exercise 14.52.** Let  $f_t$  be a family of differentiable functions at  $x_0$ . Then  $d_{x_0}f_t \in T_{x_0}^*M$  is a curve in the cotangent space. Use Proposition 7.2.1 and Exercise 7.35 to prove that

$$\left. \frac{d}{dt} \right|_{t=0} d_{x_0}f_t = d_{x_0}g \iff \left. \frac{d}{dt} \right|_{t=0} X(f_t) = X(g) \text{ for all } X,$$

and

$$\int_a^b (d_{x_0}f_t) dt = d_{x_0}g \iff \int_a^b X(f_t) dt = X(g) \text{ for all } X.$$

Moreover, prove the following possible choice of  $g$

$$\left. \frac{d}{dt} \right|_{t=0} d_{x_0}f_t = d_{x_0} \left( \left. \frac{d}{dt} \right|_{t=0} f_t \right), \quad \int_a^b (d_{x_0}f_t) dt = d_{x_0} \left( \int_a^b f_t dt \right).$$

**Exercise 14.53.** Let  $X_t \in T_{x_0}M$  be a curve in the tangent space, and let  $f_t$  be a family of differentiable functions at  $x_0$ . Prove the Leibniz rule

$$\left. \frac{d}{dt} \right|_{t=0} X_t(f_t) = \left( \left. \frac{d}{dt} \right|_{t=0} X_t \right) (f_0) + X_0 \left( \left. \frac{d}{dt} \right|_{t=0} f_t \right).$$

**Exercise 14.54.** Prove that  $T_{(x_0, y_0)}(M \times N) = T_{x_0}M \oplus T_{y_0}N$  and  $T_{(x_0, y_0)}^*(M \times N) = T_{x_0}^*M \oplus T_{y_0}^*N$ .

## Bases of Tangent and Cotangent Spaces

Let  $\sigma: U \rightarrow M$  be a chart around  $x_0 = \sigma(\vec{u}_0)$ . A function  $f$  on  $M$  gives a function  $f \circ \sigma$  on  $U$ , and a function  $g$  on  $U$  gives a function  $g \circ \sigma^{-1}$  on  $M$ . We will abuse the notation and denote  $f \circ \sigma$  by  $f$  and denote  $g \circ \sigma^{-1}$  by  $g$ . In particular, the coordinates  $u_i$  are originally functions on  $U \subset \mathbb{R}^n$ , and are also considered as functions on  $M$ .

The isomorphism  $\sigma'(\vec{u}_0): \mathbb{R}^n \rightarrow T_{x_0}M$  translates the standard basis  $\vec{e}_i \in \mathbb{R}^n$  to a basis of the tangent space

$$\partial_{u_i} = \sigma'(\vec{u}_0)(\vec{e}_i) = [\sigma(\vec{u}_0 + t\vec{e}_i)] \in T_{x_0}M.$$

By (14.3.4), the tangent vector  $\partial_{u_i}$  is simply the partial derivative

$$\partial_{u_i}(f) = \langle \partial_{u_i}, [f] \rangle = \left. \frac{d}{dt} \right|_{t=0} (f(\sigma(\vec{u}_0 + t\vec{e}_i))) = \frac{\partial f}{\partial u_i}.$$

Note that strictly speaking, the partial derivative is applied to  $f \circ \sigma$  instead of  $f$ .

The isomorphism  $(\sigma^{-1})'(x_0)^*: (\mathbb{R}^n)^* \rightarrow T_{x_0}^*M$  translates the standard dual basis  $\vec{e}_i^* = u_i \in (\mathbb{R}^n)^*$  to a basis of the cotangent space

$$du_i = [u_i \circ \sigma^{-1}] \in T_{x_0}^*M.$$

The two bases of  $T_{x_0}M$  and  $T_{x_0}^*M$  are indeed dual to each other

$$\langle \partial_{u_i}, du_j \rangle = \frac{\partial u_j}{\partial u_i} = \delta_{ij}.$$

By (7.2.2), any tangent vector  $X \in T_{x_0}M$  has the following expression in terms of the partial derivative basis

$$X = a_1 \partial_{u_1} + a_2 \partial_{u_2} + \cdots + a_n \partial_{u_n}, \quad a_i = \langle X, du_i \rangle = X(u_i). \quad (14.3.5)$$

Similarly, any cotangent vector  $\omega \in T_{x_0}^*M$  has the following expression

$$\omega = a_1 du_1 + a_2 du_2 + \cdots + a_n du_n, \quad a_i = \langle \partial_{u_i}, \omega \rangle = \omega(\partial_{u_i}). \quad (14.3.6)$$

For the special case  $\omega = df$ , by  $\langle \partial_{u_i}, df \rangle = \partial_{u_i} f = f_{u_i}$ , we have

$$df = f_{u_1} du_1 + f_{u_2} du_2 + \cdots + f_{u_n} du_n. \quad (14.3.7)$$

**Example 14.3.3.** In Example 14.3.2, for the special case  $\vec{v} = \vec{e}_i$ , we have

$$\sigma'(\vec{u}_0)(\vec{e}_i) = \sigma_{u_i}(\vec{u}_0).$$

The right side  $\sigma_{u_i}$  is the partial derivative of the map  $\sigma: \mathbb{R}^n \rightarrow \mathbb{R}^N$  in  $u_i$ , used in Sections 13.2 and 13.3 for submanifolds of Euclidean spaces. The regularity of the parameterization means that  $\sigma_{u_1}, \sigma_{u_2}, \dots, \sigma_{u_n}$  are linearly independent vectors that span the tangent space  $T_{\vec{x}_0}M$  in (8.4.1) and Example 14.3.2.

The cotangent space  $T_{\vec{x}_0}^*M$  may be identified with the linear functions on the subspace  $T_{\vec{x}_0}M \subset \mathbb{R}^N$ . Therefore  $T_{\vec{x}_0}^*M$  is spanned by the restrictions of the linear functions  $dx_1, dx_2, \dots, dx_N \in T_{\vec{x}_0}^*\mathbb{R}^N$ . Although these linear functions form a basis of  $T_{\vec{x}_0}^*\mathbb{R}^N$ , their restrictions on  $T_{\vec{x}_0}^*M$  may become linearly dependent. The cotangent vectors  $dx_1, dx_2, \dots, dx_N$  are used in Sections 13.1, 13.2, 13.3 for integrations on sub manifolds in Euclidean spaces.

**Exercise 14.55.** Given two charts at  $x_0 \in M$ , how are two corresponding bases of  $T_{x_0}M$  related? How about the two corresponding bases of  $T_{x_0}^*M$ .

**Exercise 14.56.** For a sequence of differentiable functions  $f_n$ , explain that the sequence  $d_{x_0}f_n \in T_{x_0}^*M$  converges if and only if the first order partial derivatives of  $f_n$  in a chart converge.

## Tangent and Cotangent Spaces at Boundary

Let  $\sigma: U \rightarrow M$  be a chart around a boundary point  $x_0 = \sigma(\vec{u}_0) \in \partial M$ . This means that  $U$  is an open subset of  $\mathbb{R}_+^n$  and  $\vec{u}_0 \in U \cap \mathbb{R}^{n-1} \times 0$ . We may modify the pairing (14.3.1) to

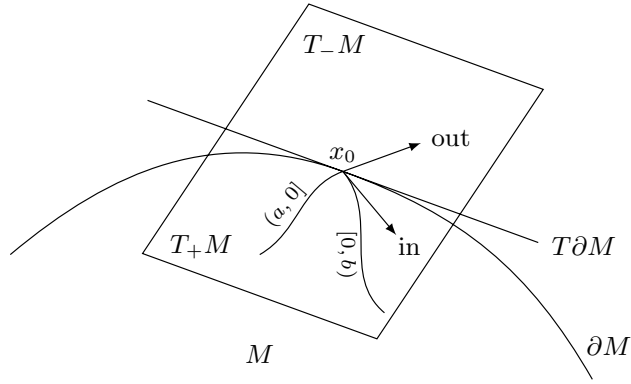
$$\langle \gamma, f \rangle_+ = \left. \frac{d}{dt} \right|_{t=0^+} f(\gamma(t)),$$

for differentiable curves  $\gamma: [0, b) \rightarrow M$  satisfying  $\gamma(0) = x_0$  and differentiable functions  $f$  at  $x_0$  (i.e.,  $f \circ \sigma$  is the restriction of a differentiable function on an open subset of  $\mathbb{R}^n$ ).

We may adopt Definition 14.3.1 to the modified pairing. Since the derivatives of functions are determined by one sided derivatives, we still get the whole cotangent vector space  $T_{x_0}^* M$ , together with the vector space structure given by  $a[f] + b[g] = [af + bg]$ . However, the equivalence classes  $[\gamma]$  form only the half tangent space  $T_{x_0+} M$ , consisting of those tangent vectors pointing toward  $M$ . Inside the half tangent space is the tangent space of the boundary manifold

$$T_{x_0} \partial M \rightarrow T_{x_0+} M: [\gamma: (a, b) \rightarrow \partial M] \mapsto [\gamma_+: [0, b) \rightarrow \partial M].$$

Exercise 14.58 shows that the map is injective, so that we can think of  $T_{x_0} \partial M$  as a subset of  $T_{x_0+} M$ .



**Figure 14.3.1.** Tangent space at boundary of manifold

The other half tangent space  $T_{x_0-} M$ , consisting of those tangent vectors pointing away from  $M$ , can be similarly defined by taking differentiable curves  $\gamma: (a, 0] \rightarrow M$  satisfying  $\gamma(0) = x_0$  and using the modified pairing

$$\langle \gamma, f \rangle_- = \left. \frac{d}{dt} \right|_{t=0-} f(\gamma(t)).$$

Adopting Definition 14.3.1 again, the equivalence classes  $[\gamma]$  form the half tangent space  $T_{x_0-} M$ , and we also have the injection

$$T_{x_0} \partial M \rightarrow T_{x_0-} M: [\gamma: (a, b) \rightarrow \partial M] \mapsto [\gamma_-: (a, 0] \rightarrow \partial M].$$

Alternatively, for any differentiable curve  $\gamma: [0, b) \rightarrow M$ , we reverse the direction and define  $\bar{\gamma}(t) = \gamma(-t): (-b, 0] \rightarrow M$ . Then we have  $\langle \bar{\gamma}, f \rangle_- = -\langle \gamma, f \rangle_+$ . This means that we may define

$$T_{x_0-} M = \{ \bar{X} : X \in T_{x_0+} M \}$$

as a copy of  $T_{x_0+} M$ , and  $T_{x_0} \partial M$  is considered as a subset of  $T_{x_0-} M$  via

$$T_{x_0} \partial M \xrightarrow{-id} T_{x_0} \partial M \rightarrow T_{x_0+} M \rightarrow T_{x_0-} M.$$

The whole tangent space  $T_{x_0}M$  may be obtained by glueing the two half tangent spaces along the subset  $T_{x_0}\partial M$ . We also note that for functions defined on open subsets of  $\mathbb{R}^n$ , we have  $\langle \gamma, f_1 \rangle_+ = \langle \gamma, f_2 \rangle_+$  for all  $\gamma: [0, b) \rightarrow M$  if and only if  $\langle \gamma, f_1 \rangle_- = \langle \gamma, f_2 \rangle_-$  for all  $\gamma: (a, 0] \rightarrow M$ . This implies that applying Definition 14.3.1 to  $\langle \gamma, f \rangle_-$  gives the same cotangent space  $T_{x_0}^*M$  as  $\langle \gamma, f \rangle_+$ .

**Proposition 14.3.4.** *The tangent and cotangent spaces  $T_{x_0}M$  and  $T_{x_0}^*M$  of a differentiable manifold  $M$  at a boundary point  $x_0 \in \partial M$  have natural vector space structures, and  $T_{x_0}M$  and  $T_{x_0}^*M$  form a dual pairing.*

*Proof.* Similar to the proof of Proposition 14.3.2, we introduce

$$(\sigma^{-1})'(x_0): T_{x_0}M \rightarrow \mathbb{R}^n, \quad \begin{cases} [[0, b) \xrightarrow{\gamma} M] & \mapsto (\sigma^{-1} \circ \gamma)'_+(0), \\ [(a, 0] \xrightarrow{\gamma} M] & \mapsto (\sigma^{-1} \circ \gamma)'_-(0); \end{cases}$$

$$\sigma'(\vec{u}_0): \mathbb{R}^n \rightarrow T_{x_0}M, \quad \vec{v} \mapsto [\sigma(\vec{u}_0 + t\vec{v})] \begin{cases} t \geq 0, & \text{if } \vec{v} \in \mathbb{R}_+^n, \\ t \leq 0, & \text{if } \vec{v} \in \mathbb{R}_-^n. \end{cases}$$

The maps for the cotangent space are the same as before. Then we verify that the maps are well defined, inverse to each other, induce vector space structure, and give dual pairing, similar to the proofs of Propositions 14.3.2 and 14.3.3.  $\square$

**Exercise 14.57.** Suppose  $U$  is an open subset of  $\mathbb{R}_+^n$  and  $\vec{u}_0 \in U \cap \mathbb{R}^{n-1} \times 0$ . Prove that the following is equivalent for a function  $g$  on  $U$ .

1.  $g$  has linear approximation on  $U$  at  $\vec{u}_0$ .
2.  $g$  is the restriction of a function on an open subset of  $\mathbb{R}^n$  that is differentiable at  $u_0$ .

**Exercise 14.58.** Prove that the map  $T_{x_0}\partial M \rightarrow T_{x_0+}M$  is well defined and injective. This means that the equivalence in  $T_{x_0}\partial M$  as tested by differentiable functions on  $\partial M$  is the same as the equivalence in  $T_{x_0+}M$  as tested by differentiable functions on  $M$ .

**Exercise 14.59.** Complete the proof of Proposition 14.3.4.

**Exercise 14.60.** Prove that  $T_{x_0}^*\partial M$  is naturally a quotient vector space of  $T_{x_0}^*M$ . Moreover, prove that the dual pairing between  $T_{x_0}\partial M$  and  $T_{x_0}^*\partial M$  is compatible with the dual pairing between  $T_{x_0}M$  and  $T_{x_0}^*M$ , with respect to the inclusion  $T_{x_0}\partial M \subset T_{x_0}M$  and the quotient  $T_{x_0}^*M \rightarrow T_{x_0}^*\partial M$ .

## 14.4 Differentiable Map

Since manifolds are locally identified with (open subsets of) Euclidean spaces, many aspects of the differentiation and integration can be carried out on manifolds. The usual modus operandi is to use charts to pull whatever happens on the manifold to the Euclidean space and then do the calculus on the Euclidean space. If we can show that this is independent of the choice of charts, then what we have done is really happening on the manifold itself. We have done this for the sequence limit, the tangent and cotangent spaces, and the dual pairing between them.

## Differentiable Map

**Definition 14.4.1.** A map  $F: M \rightarrow N$  between differentiable manifolds is *differentiable*, if  $F$  is continuous and, for any charts  $\sigma: U \rightarrow M_\sigma \subset M$  and  $\tau: V \rightarrow N_\tau \subset N$  satisfying  $F(M_\sigma) \subset N_\tau$ , the composition  $\tau^{-1} \circ F \circ \sigma$  is differentiable.

The property  $F(M_\sigma) \subset N_\tau$  is needed for the composition  $\tau^{-1} \circ F \circ \sigma$  to make sense. Proposition 14.2.13 says that we can always arrange to have  $F(M_\sigma) \subset N_\tau$  everywhere.

If  $M$  and  $N$  are  $C^r$ -manifolds and the composition  $\tau^{-1} \circ F \circ \sigma$  is  $r$ -th order continuously differentiable, then we say  $F$  is  *$r$ -th order differentiable*, or a  *$C^r$ -map*. The  $C^r$  requirement on  $M$  and  $N$  makes sure that the definition is independent of the choice of  $C^r$ -compatible charts.

**Exercise 14.61.** Prove that the differentiability is independent of the choice of charts. More specifically, prove that if  $\tau^{-1} \circ F \circ \sigma$  is differentiable for one pair of charts  $\sigma$  and  $\tau$ , then  $\bar{\tau}^{-1} \circ F \circ \bar{\sigma}$  is differentiable on the overlapping for any other pair of charts  $\bar{\sigma}$  and  $\bar{\tau}$ .

**Exercise 14.62.** Prove that composition of differentiable maps is differentiable.

**Exercise 14.63.** Prove that any chart  $\sigma: U \rightarrow M$  is differentiable, and the inverse  $\sigma^{-1}: \sigma(U) \rightarrow U$  is also differentiable.

**Exercise 14.64.** Prove that a function  $f: M \rightarrow \mathbb{R}$  is differentiable if and only if for any chart  $\sigma: U \rightarrow M$ , the composition  $f \circ \sigma$  is a differentiable function on  $U \subset \mathbb{R}^n$ .

**Exercise 14.65.** Prove that a curve  $\gamma: (a, b) \rightarrow M$  is differentiable if and only if for any  $t \in (a, b)$  and any chart  $\sigma: U \rightarrow M$  around  $\gamma(t_0)$ , there is  $\delta > 0$ , such that  $\gamma(t - \delta, t + \delta) \subset \sigma(U)$ , and  $\sigma^{-1} \circ \gamma: (t - \delta, t + \delta) \rightarrow U \subset \mathbb{R}^n$  is differentiable.

**Exercise 14.66.** Prove that under the set up in Exercise 14.43, the map  $F$  is differentiable if and only if the composition  $\tau_i^{-1} \circ F \circ \sigma_i: U_i \rightarrow V_i$  is differentiable for each  $i$ .

**Exercise 14.67.** For differentiable manifolds  $M$  and  $N$ , prove that the projection map  $M \times N \rightarrow M$  and the inclusion map  $M \rightarrow M \times y_0 \subset M \times N$  are differentiable.

For a map  $F: \mathbb{R}^m \rightarrow \mathbb{R}^n$ , the derivative  $F'(\vec{x}_0): \mathbb{R}^m \rightarrow \mathbb{R}^n$  is the linear transform in the first order term in the linear approximation of  $F$

$$F(\vec{x}) = F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x} - \vec{x}_0) + o(\vec{x} - \vec{x}_0).$$

To extend to maps between differentiable manifolds, we also need to linearly approximate the manifolds by the tangent spaces. So the derivative of a differentiable map  $F: M \rightarrow N$  at  $x_0 \in M$  should be a linear transform  $F'(x_0): T_{x_0}M \rightarrow T_{F(x_0)}N$ .

**Definition 14.4.2.** Let  $F: M \rightarrow N$  be a differentiable map between differentiable

manifolds. The *derivative* (or *pushforward*) of  $F$  is

$$F_* = F'(x_0): T_{x_0}M \rightarrow T_{F(x_0)}N, \quad [\gamma] \mapsto [F \circ \gamma].$$

The *pullback* of  $F$  is

$$F^*: T_{F(x_0)}^*N \rightarrow T_{x_0}^*M, \quad [f] \mapsto [f \circ F].$$

The derivative  $F'$  is also called pushforward because it goes in the same direction of  $F$ . Such behaviour is called *covariant*, and is often denoted by subscript  $*$ . The map  $F^*$  on cotangent vectors is called pullback because it goes in the opposite direction of  $F$ . Such behaviour is called *contravariant*, and is often denoted by superscript  $*$ . The pullback can also be applied to differentiable functions on manifolds because it is also contravariant

$$F^*: C^1(N) \rightarrow C^1(M), \quad f \mapsto f \circ F.$$

This is a linear transform and actually an algebraic homomorphism. Then the definition  $F^*[f] = [f \circ F]$  of pullback means

$$F^*df = dF^*f. \quad (14.4.1)$$

For a tangent vector  $X \in T_{x_0}M$  and a differentiable function  $f$  at  $F(x_0) \in N$ , the definition of the derivative means

$$\begin{aligned} F_*X(f) &= F'(x_0)(X)(f) = \left. \frac{d}{dt} \right|_{t=0} f(F \circ \gamma(t)) \\ &= \left. \frac{d}{dt} \right|_{t=0} (f \circ F)(\gamma(t)) = X(f \circ F) = X(F^*f). \end{aligned} \quad (14.4.2)$$

The first line happens on  $N$ , and the second line happens on  $M$ . The equality implies that  $F_* = F'(x_0)$  is well defined. For  $\omega = df$ , the equality (14.4.2) can be rephrased as

$$\langle F_*X, \omega \rangle = \langle X, F^*\omega \rangle, \quad X \in TM, \omega \in T^*N.$$

This shows exactly that  $F_*$  and  $F^*$  form a dual pair of linear transforms.

**Example 14.4.1.** A chart  $\sigma: U \rightarrow M$  around  $\vec{x}_0 = \sigma(u_0)$  is differentiable. Its inverse  $\sigma^{-1}: \sigma(U) \rightarrow U$  is also differentiable. See Exercise 14.63. Then we get derivatives (or pushforwards)

$$\sigma_* = \sigma'(\vec{u}_0): T_{\vec{u}_0}U \rightarrow T_{x_0}M, \quad (\sigma^{-1})_* = (\sigma^{-1})'(x_0): T_{x_0}M \rightarrow T_{\vec{u}_0}U,$$

and pullbacks

$$\sigma^* = \sigma'(\vec{u}_0)^*: T_{x_0}^*M \rightarrow T_{\vec{u}_0}^*U, \quad (\sigma^{-1})^* = (\sigma^{-1})'(x_0)^*: T_{\vec{u}_0}^*U \rightarrow T_{x_0}^*M.$$

Combined with the isomorphisms  $T_{\vec{u}_0}U \cong \mathbb{R}^n$  and  $T_{\vec{u}_0}^*U \cong (\mathbb{R}^n)^*$  in Example 14.3.1, we get

$$\begin{aligned} \sigma'(\vec{u}_0): \mathbb{R}^n &\cong T_{\vec{u}_0}U \rightarrow T_{x_0}M, & \vec{v} &\mapsto [\vec{u}_0 + t\vec{v}] \mapsto [\sigma(\vec{u}_0 + t\vec{v})]; \\ (\sigma^{-1})'(x_0): T_{x_0}M &\rightarrow T_{\vec{u}_0}U \cong \mathbb{R}^n, & [\gamma] &\mapsto [\sigma^{-1} \circ \gamma] \mapsto (\sigma^{-1} \circ \gamma)'(0); \\ \sigma'(\vec{u}_0)^*: T_{x_0}^*M &\rightarrow T_{\vec{u}_0}^*U \cong (\mathbb{R}^n)^*, & [f] &\mapsto [f \circ \sigma] \mapsto (f \circ \sigma)'(\vec{u}_0); \\ (\sigma^{-1})'(x_0)^*: (\mathbb{R}^n)^* &\cong T_{\vec{u}_0}^*U \rightarrow T_{x_0}^*M, & l &\mapsto [l] \mapsto [l \circ \sigma^{-1}]. \end{aligned}$$

This explains the notations for the isomorphisms in Proposition 14.3.2.

**Example 14.4.2.** A function  $f: M \rightarrow \mathbb{R}$  is differentiable if and only if for any chart  $\sigma: U \rightarrow M$ , the composition  $f \circ \sigma$  is differentiable. See Exercise 14.64. The derivative of the function is a linear functional on the tangent space

$$f_*[\gamma] = f'(x_0)([\gamma]) = [f \circ \gamma] = (f \circ \gamma)'(0): T_{x_0}M \rightarrow T_{f(x_0)}\mathbb{R} \cong \mathbb{R}.$$

Here the second equality is the definition of the derivative  $f'(x_0): T_{x_0}M \rightarrow T_{f(x_0)}\mathbb{R}$ , and the third equality is the isomorphism  $T_{f(x_0)}\mathbb{R} \cong \mathbb{R}$  in Example 14.3.1. By (14.3.2),  $(f \circ \gamma)'(0)$  is simply the pairing between the tangent vector  $X = [\gamma]$  and the cotangent vector  $df = [f]$ . Therefore the linear functional  $f_* = f'(x_0)$  is exactly the differential  $f'(x_0) = d_{x_0}f \in T_{x_0}^*M$ .

Alternatively, the differentiable function induces a pullback  $f^*: T^*\mathbb{R} \rightarrow T_{x_0}^*M$ . The standard basis in  $T^*\mathbb{R}$  is the differential  $dt$  of the identity linear functional  $t \mapsto t: \mathbb{R} \rightarrow \mathbb{R}$  (the standard basis of  $\mathbb{R}^*$ ). The pullback  $f^*$  is determined by  $f^*dt = [(t \mapsto t) \circ f] = [f] = df$ . Therefore the pullback  $f^*$  is naturally identified with its differential  $df \in T_{x_0}^*M$ .

**Example 14.4.3.** Let  $\gamma: (a, b) \rightarrow M$  be a differentiable curve. See Exercise 14.65. The derivative linear transform

$$\gamma'(t_0): \mathbb{R} = T_{t_0}(a, b) \mapsto T_{\gamma(t_0)}M$$

is determined by its value  $\gamma'(t_0)(1) \in T_{\gamma(t_0)}M$  at  $1 \in \mathbb{R}$ . We abuse the notation by simply also use  $\gamma'(t_0)$  to denote the *tangent vector*  $\gamma'(t_0)(1)$  of the curve.

Note that  $1 \in \mathbb{R}$  corresponds to the equivalence class  $[t_0 + t \cdot 1] \in T_{t_0}(a, b)$ . Therefore the tangent vector of the curve

$$\gamma'(t_0) = \gamma'(t_0)(1) = [\gamma(t_0 + t)]_{\text{at } t=0} = [\gamma(t)]_{\text{at } t=t_0} \in T_{\gamma(t_0)}M$$

is exactly the equivalence class represented by the curve itself.

**Exercise 14.68.** Prove that the pullback of cotangent vectors is well defined. In other words,  $[f \circ F]$  depends only on the equivalence class  $[f]$ .

**Exercise 14.69.** Formulate and prove the *chain rules* for the derivative and the pullback.

**Exercise 14.70.** Extend the discussion in Example 14.4.2 to vector valued functions  $f: M \rightarrow \mathbb{R}^m$ .

**Exercise 14.71.** What is the pullback of a continuously differentiable curve  $\gamma: (a, b) \rightarrow M$ ?

**Exercise 14.72.** What is the derivative and the pullback of the diagonal map  $F(x) = (x, x): M \rightarrow M \times M$ ?

**Exercise 14.73.** Describe the derivatives and the pullbacks of the projection map  $M \times N \rightarrow M$  and the inclusion map  $M \rightarrow M \times y_0 \subset M \times N$ .

**Exercise 14.74.** A submanifold  $M$  of  $\mathbb{R}^N$  has a natural inclusion map  $i: M \rightarrow \mathbb{R}^N$ . Prove that the tangent space defined in Section 8.4 is the image of the injective derivative map  $i'(\vec{x}_0): T_{\vec{x}_0}M \rightarrow T_{\vec{x}_0}\mathbb{R}^N \cong \mathbb{R}^N$ .

**Exercise 14.75.** Prove that the derivative of the map  $F(x, y, z) = (x^2 - y^2, xy, xz, yz): \mathbb{R}^3 \rightarrow \mathbb{R}^4$  at  $(x, y, z)$  is injective as long as  $x \neq 0$  or  $y \neq 0$ . Then prove that the derivative of the restriction  $\bar{F}: S^2 \rightarrow \mathbb{R}^4$  of  $F$  is always injective.

## Differentiation in Charts

Let  $\sigma$  and  $\tau$  be charts satisfying  $F(\sigma(U)) \subset \tau(V)$ , and  $x_0 = \sigma(\vec{u}_0)$ ,  $F(x_0) = \tau(\vec{v}_0)$ . Then the map  $\tau^{-1} \circ F \circ \sigma: U \subset \mathbb{R}^m \rightarrow V \subset \mathbb{R}^n$  is explicit expressions of the coordinates  $(v_1, v_2, \dots, v_n)$  of  $V$  as functions of the coordinates  $(u_1, u_2, \dots, u_m)$  of  $U$ . The tangent spaces  $T_{x_0}M$  and  $T_{F(x_0)}N$  have bases  $\{\partial_{u_1}, \partial_{u_2}, \dots, \partial_{u_m}\}$  and  $\{\partial_{v_1}, \partial_{v_2}, \dots, \partial_{v_n}\}$ , and we have

$$F_*\partial_{u_i} = a_{i1}\partial_{v_1} + a_{i2}\partial_{v_2} + \cdots + a_{in}\partial_{v_n}.$$

By (14.3.5), the coefficients are

$$a_{ij} = F_*\partial_{u_i}(v_j) = \partial_{u_i}(v_j \circ F) = \frac{\partial v_j}{\partial u_i}.$$

Therefore

$$\begin{aligned} F_*\partial_{u_1} &= \frac{\partial v_1}{\partial u_1}\partial_{v_1} + \frac{\partial v_2}{\partial u_1}\partial_{v_2} + \cdots + \frac{\partial v_n}{\partial u_1}\partial_{v_n}, \\ F_*\partial_{u_2} &= \frac{\partial v_1}{\partial u_2}\partial_{v_1} + \frac{\partial v_2}{\partial u_2}\partial_{v_2} + \cdots + \frac{\partial v_n}{\partial u_2}\partial_{v_n}, \\ &\vdots \\ F_*\partial_{u_m} &= \frac{\partial v_1}{\partial u_m}\partial_{v_1} + \frac{\partial v_2}{\partial u_m}\partial_{v_2} + \cdots + \frac{\partial v_n}{\partial u_m}\partial_{v_n}. \end{aligned}$$

The coefficients form the Jacobian matrix of the map  $\tau^{-1} \circ F \circ \sigma$ .

**Example 14.4.4.** In Example 14.2.5, we find the explicit formula for the canonical map  $F(\vec{x}) = [\vec{x}]: S^n \rightarrow \mathbb{R}P^n$

$$\tau_0^{-1} \circ F \circ \sigma_0(\vec{u}) = \frac{\vec{u}}{\sqrt{1 - \|\vec{u}\|_2^2}}: U \rightarrow V$$

in terms of two charts of  $S^n$  and  $\mathbb{R}P^n$ . Taking  $\vec{u} = \vec{u}_0 + t\vec{v}$ , we get

$$\begin{aligned} \tau_0^{-1} \circ F \circ \sigma_0(\vec{u}) &= \frac{\vec{u}_0 + t\vec{v}}{\sqrt{1 - \|\vec{u}_0 + t\vec{v}\|_2^2}} \\ &= (\vec{u}_0 + t\vec{v}) [1 - \|\vec{u}_0\|^2 - 2(\vec{u}_0 \cdot \vec{v})t + o(t)]^{-\frac{1}{2}} \\ &= (\vec{u}_0 + t\vec{v}) (1 - \|\vec{u}_0\|^2)^{-\frac{1}{2}} \left[ 1 - 2\frac{\vec{u}_0 \cdot \vec{v}}{1 - \|\vec{u}_0\|^2}t + o(t) \right]^{-\frac{1}{2}} \\ &= \frac{\vec{u}_0}{\sqrt{1 - \|\vec{u}_0\|_2^2}} + \frac{1}{\sqrt{1 - \|\vec{u}_0\|_2^2}} \left[ \vec{v} + \frac{\vec{u}_0 \cdot \vec{v}}{1 - \|\vec{u}_0\|^2} \vec{u}_0 \right] t + o(t). \end{aligned}$$



Therefore the explicit formula for the derivative is

$$(\tau_0^{-1} \circ F \circ \sigma_0)'(\vec{u}_0)(\vec{v}) = \frac{1}{\sqrt{1 - \|\vec{u}_0\|_2^2}} \left[ \vec{v} + \frac{\vec{u}_0 \cdot \vec{v}}{1 - \|\vec{u}_0\|_2^2} \vec{u}_0 \right].$$

The derivative is the identity at  $\vec{u}_0 = \vec{0}$ , and is the following at  $\vec{u}_0 \neq \vec{0}$

$$(\tau_0^{-1} \circ F \circ \sigma_0)'(\vec{u}_0)(\vec{v}) = \begin{cases} \frac{\vec{v}}{\sqrt{1 - \|\vec{u}_0\|_2^2}}, & \text{if } \vec{v} \perp \vec{u}_0, \\ \frac{\vec{v}}{(\sqrt{1 - \|\vec{u}_0\|_2^2})^3}, & \text{if } \vec{v} \parallel \vec{u}_0. \end{cases}$$

Similar charts  $\sigma_i$  and  $\tau_i$  can be found for  $x_i > 0$  and for  $x_i \neq 0$ , as well as for  $x_i < 0$  and for  $x_i \neq 0$ . The differentiations can also be calculated.

**Example 14.4.5.** Let  $F(\xi) = 2\xi: S_{\frac{1}{2}}^n \rightarrow S^n$  be the scaling map from the sphere of radius  $\frac{1}{2}$  to the unit sphere. The stereographic projections in Example 14.1.3 give an atlas for  $S_{\frac{1}{2}}^n$ . Let  $\xi = (\eta, z)$ , with  $\eta \in \mathbb{R}^n$  and  $z \in \mathbb{R}$ . Since the triangles  $N\xi\vec{v}$  and  $S\xi\vec{u}$  are similar at the ratio of  $\|\vec{v}\| : \|\vec{u}\|$ , we get

$$\frac{\|\eta\|}{\|\vec{u}\|} = \frac{\|\vec{v}\|}{\|\vec{u}\| + \|\vec{v}\|}, \quad z = \frac{1}{2} \frac{\|\vec{u}\|}{\|\vec{u}\| + \|\vec{v}\|} - \frac{1}{2} \frac{\|\vec{v}\|}{\|\vec{u}\| + \|\vec{v}\|} = \frac{1}{2} \frac{\|\vec{u}\| - \|\vec{v}\|}{\|\vec{u}\| + \|\vec{v}\|}.$$

Then by  $\eta$  parallel to  $\vec{v}$  and  $\|\vec{u}\|\|\vec{v}\| = 1$ , we get

$$\eta = \frac{\|\vec{u}\|\|\vec{v}\|}{\|\vec{u}\| + \|\vec{v}\|} \frac{\vec{v}}{\|\vec{v}\|} = \frac{\vec{v}}{1 + \|\vec{v}\|^2}, \quad z = \frac{1}{2} \frac{1 - \|\vec{v}\|^2}{1 + \|\vec{v}\|^2}, \quad 2\xi = \frac{(2\vec{v}, 1 - \|\vec{v}\|^2)}{1 + \|\vec{v}\|^2}.$$

On the other hand, the upper half unit sphere has a chart

$$\sigma_0(\vec{w}) = \left( \vec{w}, \sqrt{1 - \|\vec{w}\|^2} \right) : \{ \vec{w} \in \mathbb{R}^n : \|\vec{w}\| < 1 \} \rightarrow S^n.$$

Then we have

$$\sigma_0^{-1} \circ F \circ \sigma_S(\vec{v}) = \frac{2\vec{v}}{1 + \|\vec{v}\|^2} : \{ \vec{v} \in \mathbb{R}^n : \|\vec{v}\| < 1 \} \rightarrow \{ \vec{w} \in \mathbb{R}^n : \|\vec{w}\| < 1 \}.$$

Taking  $\vec{v} = \vec{v}_0 + t\vec{a}$ , we get

$$\begin{aligned} \sigma_0^{-1} \circ F \circ \sigma_S(\vec{v}) &= \frac{2(\vec{v}_0 + t\vec{a})}{1 + \|\vec{v}_0 + t\vec{a}\|^2} \\ &= \frac{2(\vec{v}_0 + t\vec{a})}{1 + \|\vec{v}_0\|^2 + 2(\vec{v}_0 \cdot \vec{a})t + o(t)} \\ &= \frac{2(\vec{v}_0 + t\vec{a})}{1 + \|\vec{v}_0\|^2} \left[ 1 - 2 \frac{\vec{v}_0 \cdot \vec{a}}{1 + \|\vec{v}_0\|^2} t + o(t) \right] \\ &= \frac{2\vec{v}_0}{1 + \|\vec{v}_0\|^2} + \frac{2}{1 + \|\vec{v}_0\|^2} \left[ \vec{a} - 2 \frac{\vec{v}_0 \cdot \vec{a}}{1 + \|\vec{v}_0\|^2} \vec{v}_0 \right] t + o(t). \end{aligned}$$

The derivative is

$$(\sigma_0^{-1} \circ F \circ \sigma_S)'(\vec{v}_0)(\vec{a}) = \begin{cases} \frac{2}{1 + \|\vec{v}_0\|^2} \vec{a}, & \text{if } \vec{a} \perp \vec{v}_0, \\ \frac{2(1 - \|\vec{v}_0\|^2)}{(1 + \|\vec{v}_0\|^2)^2} \vec{a}, & \text{if } \vec{a} \parallel \vec{v}_0. \end{cases}$$

**Exercise 14.76.** Given charts of  $M$  and  $N$ , the cotangent spaces  $T^*M$  and  $T^*N$  have bases  $du_i$  and  $dv_j$ . Find the explicit formula for the pullback in terms of the bases.

**Exercise 14.77.** Show that the obvious map  $M \rightarrow S^1: (z, v) \mapsto z$  from the open Möbius band in Example 14.1.4 to the circle is differentiable. Moreover, construct the obvious inclusion map  $S^1 \rightarrow M$  and show that it is differentiable.

**Exercise 14.78.** For the Möbius band in Example 14.1.9, construct the obvious map  $M \rightarrow S^1$  and two maps  $S^1 \rightarrow M$ , one being similar to the one in Exercise 14.77, and the other being the boundary of the Möbius band. Show that all three maps are differentiable.

**Exercise 14.79.** Show that the canonical map  $S^{2n+1} \rightarrow \mathbb{C}P^n$  is differentiable and find the derivative in terms of suitable charts.

**Exercise 14.80.** By thinking of real numbers as complex numbers, we have the map  $\mathbb{R}P^n \rightarrow \mathbb{C}P^n$ . Show that the map is injective and differentiable.

## Diffeomorphism

**Definition 14.4.3.** A *diffeomorphism* between differentiable manifolds is an invertible map  $F: M \rightarrow N$ , such that both  $F$  and  $F^{-1}$  are differentiable.

Given the diffeomorphism  $F$ , whatever differentiation or integration on  $M$  can be translated by  $F$  to differentiation or integration on  $N$ , and vice versa. In particular, calculus on  $M$  is equivalent to calculus on  $N$ .

**Example 14.4.6.** For a continuously differentiable map  $F: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^{N-n}$ , its graph

$$\Gamma_F = \{(\vec{x}, F(\vec{x})): \vec{x} \in U\} \subset \mathbb{R}^N$$

is a manifold. The maps

$$U \rightarrow \Gamma_F: \vec{x} \mapsto (\vec{x}, F(\vec{x})), \quad \Gamma_F \rightarrow U: (\vec{x}, \vec{y}) \mapsto \vec{x}$$

are inverse to each other and are continuously differentiable. Therefore  $\Gamma_F$  and  $U$  are diffeomorphic.

In general, Exercise 14.63 says that if  $\sigma: U \rightarrow M$  is a chart covering the whole manifold  $M$ :  $\sigma(U) = M$ , then  $\sigma$  is a diffeomorphism between  $U$  and  $M$ .

**Example 14.4.7.** The sphere of half radius in Example 14.1.3 and the complex projective space  $\mathbb{C}P^1$  in Example 14.1.6 are covered by two charts, with transition maps

$$\varphi_{NS}(u, v) = \frac{(u, v)}{u^2 + v^2}, \quad \varphi_{10}(u, v) = \frac{(u, -v)}{u^2 + v^2}: \mathbb{R}^2 - (0, 0) \rightarrow \mathbb{R}^2 - (0, 0).$$

The transitions become the same if we modify  $\sigma_N(u, v)$  to  $\sigma_N(u, -v)$ . This suggests that  $S^2$  and  $\mathbb{C}P^1$  are diffeomorphic. See Exercise 14.8.

Specifically, denote vectors  $(u, v) \in \mathbb{R}^2$  by complex numbers  $w = u + iv \in \mathbb{C}$ . Then the modification  $(u, v) \mapsto (u, -v)$  means taking the complex conjugation  $\bar{w}$  of  $w$ . According

to the calculation in Example 14.4.5, the sphere  $S^2 \subset \mathbb{C} \times \mathbb{R}$  of unit radius is given by the following modified stereographic projection charts

$$\tilde{\sigma}_S(w) = 2\sigma_S(w) = \frac{(2w, 1 - |w|^2)}{1 + |w|^2}, \quad \tilde{\sigma}_N(w) = 2\sigma_N(\bar{w}) = \frac{(2\bar{w}, |w|^2 - 1)}{|w|^2 + 1}.$$

Then the transition  $\tilde{\varphi}_{NS} = \tilde{\sigma}_N^{-1} \circ \tilde{\sigma}_S(w) = \frac{1}{w}$  is equal to the transition  $\varphi_{10}(w) = \frac{1}{w}$ , and the diffeomorphism  $F: \mathbb{CP}^1 \rightarrow S^2$  should be given by  $F(\sigma_0(w)) = \tilde{\sigma}_S(w)$  and  $F(\sigma_1(w)) = \tilde{\sigma}_N(\bar{w})$ . This means

$$F([1, w]) = \frac{(2w, 1 - |w|^2)}{1 + |w|^2}, \quad F([w, 1]) = \frac{(2\bar{w}, |w|^2 - 1)}{|w|^2 + 1},$$

or

$$F([w_0, w_1]) = \frac{(2\bar{w}_0 w_1, |w_0|^2 - |w_1|^2)}{|w_0|^2 + |w_1|^2}: \mathbb{CP}^1 \rightarrow S^2 \subset \mathbb{C} \times \mathbb{R}.$$

By  $(\tilde{\sigma}_S^{-1} \circ F \circ \sigma_0)(w) = w$  and  $(\tilde{\sigma}_N^{-1} \circ F \circ \sigma_1)(w) = w$ , both  $F$  and  $F^{-1}$  are differentiable. Roughly speaking, the identification of  $S^2$  and  $\mathbb{CP}^1$  is given by

$$S^2 \cong \mathbb{C} \cup \infty \rightarrow \mathbb{CP}^1: w \mapsto [w, 1], \quad \infty \mapsto [1, 0],$$

and

$$\mathbb{CP}^1 \rightarrow \mathbb{C} \cup \infty \cong S^2: [w_0, w_1] \mapsto \frac{w_1}{w_0}, \quad [0, w_1] \mapsto \infty.$$

Here  $\infty$  is the south pole, and  $\mathbb{C}$  is regarded as a subset of  $S^2$  through the stereographic projection  $\tilde{\sigma}_S$ . The formulae we derived above make the idea rigorous.

**Exercise 14.81.** Show that the circle  $S^1$  and the real projective space  $\mathbb{RP}^1$  are diffeomorphic. See Exercise 14.7.

## Local Diffeomorphism

**Definition 14.4.4.** A differentiable map  $F: M \rightarrow N$  is a *local diffeomorphism* if for any  $F(x) = y$ , there is an open neighbourhood  $B$  of  $x$ , such that  $F(B)$  is an open neighbourhood of  $y$ , and  $F: B \rightarrow F(B)$  is a diffeomorphism.

The Inverse Function Theorem (Theorem 8.3.1) says that the invertibility of a map between open subsets of  $\mathbb{R}^n$  can be locally detected by the invertibility of its linear approximation. By using the charts to translate maps between manifolds to this special case, we get the manifold version of the Inverse Function Theorem.

**Theorem 14.4.5 (Inverse Function Theorem).** *A continuously differentiable map between differentiable manifolds is a local diffeomorphism if and only if its derivative is invertible everywhere.*

**Example 14.4.8.** The map  $F(\theta) = (\cos \theta, \sin \theta): \mathbb{R} \rightarrow S^1$  is not a diffeomorphism because it sends infinitely many  $\theta$  to the same point on the circle. However, if  $0 < \delta < \pi$ , then for any  $\theta$ , the map  $F: (\theta - \delta, \theta + \delta) \rightarrow S^1$  is a diffeomorphism. Therefore  $F$  is a local diffeomorphism.

**Example 14.4.9.** In Example 14.2.5, we find the explicit formula for the canonical map  $F(\vec{x}) = [\vec{x}]: S^n \rightarrow \mathbb{R}P^n$  in terms of one pair of charts. In Example 14.4.4, we further calculated the derivative of the explicit formula and find that the derivative is invertible. Therefore the map is a local diffeomorphism at least on the related chart. Similar computation shows that the map is a local diffeomorphism on all the other charts. Therefore  $F$  is a local homeomorphism.

In fact,  $F$  is a two-to-one map because two unit length vectors  $\vec{x}$  and  $\vec{y}$  span the same 1-dimensional subspace if and only if  $\vec{x} = \pm\vec{y}$ . If we restrict to the chart  $\sigma_0$ , which is the half sphere with  $x_0 > 0$ , then the explicit formula for  $\tau_0^{-1} \circ F \circ \sigma_0$  in Example 14.4.4 is a full diffeomorphism between the charts  $\sigma_0$  and  $\tau_0$ . This explicitly shows that  $F$  is a local diffeomorphism on  $\sigma_0$ . The same happens to the other charts.

**Exercise 14.82.** Prove that the map  $F: S^1 \rightarrow S^1$  that sends angle  $\theta$  to  $5\theta$  is a local diffeomorphism. What about sending  $\theta$  to  $\pi - 3\theta$ ?

**Exercise 14.83.** Construct a local diffeomorphism from the open Möbius band in Example 14.1.4 to itself, such that  $z(\theta)$  goes to  $z(3\theta)$ . Moreover, construct a local diffeomorphism from the cylinder  $S^1 \times \mathbb{R}$  to the open Möbius band, such that  $z(\theta)$  goes to  $z(2\theta)$ .

**Exercise 14.84.** Prove that an invertible local diffeomorphism is a diffeomorphism.

**Exercise 14.85.** The map  $\bar{F}: S^2 \rightarrow \mathbb{R}^4$  in Exercise 14.75 satisfies  $\bar{F}(-\vec{x}) = \bar{F}(\vec{x})$  and therefore induces a map  $G[\vec{x}] = \bar{F}(\vec{x}): \mathbb{R}P^2 \rightarrow \mathbb{R}^4$ . Prove that the derivative of  $G$  is always injective.

## 14.5 Orientation

A differentiable manifold has a tangent vector space at every point. An orientation of the manifold is a compatible choice of one orientation for each tangent space. The compatibility comes from the following isomorphism between tangent spaces at nearby points. For any  $x_0$ , let  $\sigma: U \rightarrow M$  be a chart around  $x_0 = \sigma(\vec{u}_0)$ . Then for each  $x = \sigma(\vec{u})$ , we have isomorphism

$$\sigma'(\vec{u}) \circ \sigma'(\vec{u}_0)^{-1}: T_{x_0}M \xrightarrow[\cong]{\sigma'(\vec{u}_0)} \mathbb{R}^n \xrightarrow[\cong]{\sigma'(\vec{u})} T_xM. \quad (14.5.1)$$

The compatibility means that the isomorphism translates the orientation of  $T_{x_0}M$  to the orientation of  $T_xM$ .

**Definition 14.5.1.** An *orientation* of a differentiable manifold  $M$  is a choice of an orientation  $o_x$  of the tangent space  $T_xM$  for each  $x \in M$  satisfying the following compatibility condition: For any  $x_0$ , there is a chart  $\sigma: U \rightarrow M$  around  $x_0$ , such that for every  $x = \sigma(\vec{u})$ , the isomorphism (14.5.1) translates the orientation  $o_{x_0}$  to the orientation  $o_x$ . A differentiable manifold is *orientable* if it has an orientation. Otherwise it is *non-orientable*.

The compatibility condition should be independent of charts. Suppose  $\tau: V \rightarrow M$  is another chart containing  $x_0 = \tau(\vec{v}_0)$ . By shrinking  $\sigma(U)$  and  $\tau(V)$  (i.e.,

restricting  $\sigma$  and  $\tau$  to smaller open subsets) to the connected component of  $\sigma(U) \cap \tau(V)$  containing  $x_0$ , we may assume that  $\sigma(U) = \tau(V)$ , and  $U$  and  $V$  are connected. Now we compare the isomorphisms (14.5.1) for the two charts

$$T_x M \xleftarrow[\cong]{\tau'(\vec{v})} \mathbb{R}^n \xrightarrow[\cong]{\tau'(\vec{v}_0)} T_{x_0} M \xleftarrow[\cong]{\sigma'(\vec{u}_0)} \mathbb{R}^n \xrightarrow[\cong]{\sigma'(\vec{u})} T_x M.$$

Let  $\varphi = \sigma^{-1} \circ \tau$  be the transition between the two charts. By Exercise 7.60, we get

$$\begin{aligned} & \det(\sigma'(\vec{u}) \circ \sigma'(\vec{u}_0)^{-1} \circ \tau'(\vec{v}_0) \circ \tau'(\vec{v})^{-1}) \\ &= \det(\sigma'(\vec{u}_0)^{-1} \circ \tau'(\vec{v}_0) \circ \tau'(\vec{v})^{-1} \circ \sigma'(\vec{u})) \\ &= \det(\varphi'(\vec{u}_0) \circ \varphi'(\vec{u})^{-1}) = (\det \varphi'(\vec{u}_0))(\det \varphi'(\vec{u}))^{-1}. \end{aligned}$$

Since this is a continuous non-vanishing function on connected  $U$ , and takes value 1 at  $\vec{u} = \vec{u}_0$ , by Proposition 6.2.4, it is positive throughout  $U$ . This implies that if  $\sigma$  translates  $o_{x_0}$  to  $o_x$ , then  $\tau$  also translates  $o_{x_0}$  to  $o_x$ . Therefore the translation of orientation at  $x_0$  to nearby  $x$  is independent of the charts, as long as  $x$  lies in the connected component of the overlapping containing  $x_0$ .

**Exercise 14.86.** Suppose three points  $x, y, z$  are contained in a chart  $\sigma$ , and  $o_x, o_y, o_z$  are orientations at the three points. Prove that if  $\sigma$  translates  $o_x$  to  $o_y$  and  $o_y$  to  $o_z$ , then  $\sigma$  translates  $o_x$  to  $o_z$ .

**Exercise 14.87.** Prove that if a chart translates  $o_x$  to  $o_y$ , then it translates  $-o_x$  to  $-o_y$ , and translates  $o_y$  to  $o_x$ .

**Exercise 14.88.** Suppose  $M$  is an oriented manifold. Prove that the choices  $-o_x$  also give an orientation of  $M$ . We usually denote the manifold with  $-o_x$  orientation by  $-M$ .

**Exercise 14.89.** Prove that any open subset of an orientable manifold is orientable.

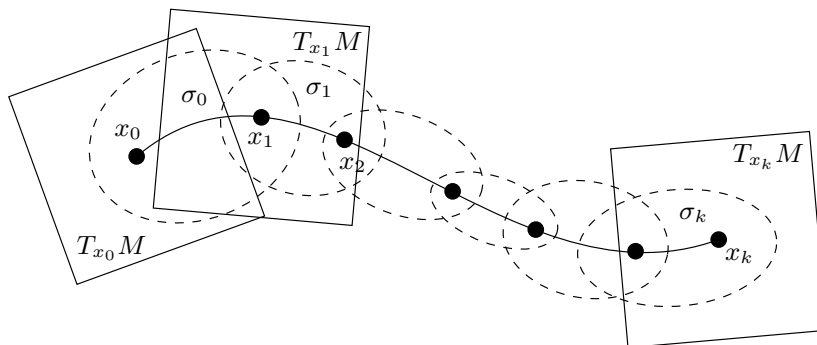
**Exercise 14.90.** Suppose  $M$  and  $N$  are oriented. Construct the product orientation on  $M \times N$  by using Exercises 7.81 and 14.54, and then prove that  $M \times N$  is also oriented. Conversely, if  $M \times N$  is orientable, must  $M$  and  $N$  be orientable?

**Exercise 14.91.** Prove that if  $M$  and  $N$  are oriented manifolds, then the product orientation satisfies  $M \times N = (-1)^{mn} N \times M$ ,  $m = \dim M$ ,  $n = \dim N$ .

## Translation of Orientation Along Curve

Suppose  $o_{x_0}$  is an orientation at one point  $x_0 \in M$ . Suppose  $\gamma: [0, 1] \rightarrow M$  is a continuous path starting at  $\gamma(0) = x_0$ . For each  $t \in [0, 1]$ , there is an interval  $(t - \delta_t, t + \delta_t)$  and a chart  $\sigma_t: U_t \rightarrow M$ , such that  $\gamma(t - \delta_t, t + \delta_t) \subset \sigma(U_t)$ . Here,  $\gamma(t) = \gamma(0)$  for  $t < 0$  and  $\gamma(t) = \gamma(1)$  for  $t > 1$ . The bounded closed interval  $[0, 1]$  is covered by open intervals  $(t - \delta_t, t + \delta_t)$ ,  $t \in [0, 1]$ . By Heine-Borel Theorem (Theorem 1.5.6) and changing notations, we get  $[0, 1] \subset (a_0, b_0) \cup (a_1, b_1) \cup \cdots \cup (a_k, b_k)$ , such that each segment  $\gamma(a_i, b_i)$  is contained in a chart  $\sigma_i: U_i \rightarrow M$ . Moreover, we may

assume the intervals are arranged in “ascending order”, in the sense that there is a partition  $1 = t_0 < t_1 < \cdots < t_k = b$ , such that  $t_i \in (a_{i-1}, b_{i-1}) \cap (a_i, b_i)$ .



**Figure 14.5.1.** *Translation of orientation along a curve*

We may use  $\sigma_0$  to translate the orientation  $o_{x_0}$  at  $x_0 = \gamma(t_0)$  to an orientation  $o_{x_1}$  at  $x_1 = \gamma(t_1)$ . Then may use  $\sigma_1$  to translate the orientation  $o_{x_1}$  to an orientation  $o_{x_2}$  at  $x_2 = \gamma(t_2)$ . Keep going, we get an orientation  $o_{x_k}$  at  $x_k = \gamma(1)$  at the end.

To see that the translation from  $o_{x_0}$  to  $o_{x_k}$  is unique, we compare two partitions and the corresponding sequences of charts. By using the common refinement, we may assume that one partition is a refinement of another. Then the problem is reduced to comparing the translation of an orientation at  $x_{i-1}$  to an orientation at  $x_i$  by using one chart  $\sigma_i$ , and the translation by following a sequence  $x_{i-1} = y_0, y_1, \dots, y_l = x_i$  and using a sequence of charts  $\tau_0, \tau_1, \dots, \tau_l$  from the refinement. Since the curve segment between  $y_{j-1}$  and  $y_j$  is connected and is therefore contained in the connected component of the overlapping between  $\sigma_i$  and  $\tau_j$ , the translation of an orientation from  $y_{j-1}$  to  $y_j$  by using  $\tau_j$  is the same as the translation by using  $\sigma_i$ . Therefore we may assume that all  $\tau_j$  are equal to  $\sigma_i$ . Then by Exercise 14.86, it is easy to see that the two translations are the same.

**Exercise 14.92.** Extend Exercises 14.86 and 14.87 to the translation along a curve.

Suppose  $M$  is a connected manifold, and an orientation  $o_{x_0}$  is fixed. For any  $x \in M$ , there is a curve  $\gamma$  connecting  $x_0$  to  $x$ , and we may translate  $o_{x_0}$  along  $\gamma$  to an orientation  $o_x$  at  $x$ . Like the integral of 1-form along a curve, the problem is whether the translation depends on the choice of  $\gamma$ . If the translation does not depend on the choice, then we get a well defined orientation  $o_x$  at each point  $x \in M$ . The use of translation in constructing the orientations  $o_x$  implies that the compatibility condition is satisfied. Therefore the orientation  $o_{x_0}$  at one point determines an orientation of  $M$ .

In general, a manifold is orientable if and only if each connected component is orientable. Moreover, an orientation is determined by the choice of an orientation at one point of each connected component.

It is easy to see that the translation along two curves give different orientations at the end if and only if the translation of an orientation  $o$  along a loop (a curve

starting and ending at the same point) gives the negative orientation  $-o$ . This is equivalent to the non-orientability of the manifold, and is also equivalent to that the manifold contains  $(\text{Möbius band}) \times \mathbb{R}^{n-1}$  as an open subset.

### Orientation Compatible Atlas

Suppose  $M$  is an oriented manifold. A chart  $\sigma: U \rightarrow M$  is *orientation compatible* if the isomorphism  $\sigma'(\vec{u}): \mathbb{R}^n \rightarrow T_x M$  always translates the standard orientation of  $\mathbb{R}^n$  to the orientation  $o_x$  of  $T_x M$ . In other words,  $o_x$  is given by the standard basis  $\{\partial_{u_1}, \partial_{u_2}, \dots, \partial_{u_n}\}$  of  $T_x M$ .

An atlas is orientation compatible if all its charts are orientation compatible. Suppose  $\sigma$  and  $\tau$  are two charts in such an atlas, and  $x = \sigma(\vec{u}) = \tau(\vec{v})$  is a point in the overlapping. Then both isomorphisms

$$\sigma'(\vec{u}): \mathbb{R}^n \cong T_x M, \quad \tau'(\vec{v}): \mathbb{R}^n \cong T_x M,$$

translate the standard orientation of  $\mathbb{R}^n$  to the same orientation  $o_x$  of  $T_x M$ . This implies that the composition  $\tau'(\vec{v})^{-1} \circ \sigma'(\vec{u}) = \varphi'(\vec{u})$  preserves the orientation of  $T_x M$ . This means that the derivative of the transition map has positive determinant. The property is used for defining the integral of 2-forms in (13.2.6).

**Proposition 14.5.2.** *A differentiable manifold is orientable if and only if there is an atlas, such that the determinants of the derivatives of the transition maps are all positive.*

*Proof.* Suppose  $M$  has an orientation given by  $o_x$  for every  $x \in M$ . Then for any  $x_0$ , there is a chart  $\sigma: U \rightarrow \mathbb{R}^n$  containing  $x_0 = \sigma(\vec{u}_0)$ , such that the isomorphism (14.5.1) translates  $o_{x_0}$  to  $o_x$  for any  $x = \sigma(\vec{u})$ ,  $\vec{u} \in U$ . If  $\sigma'(\vec{u}_0)$  translates the standard orientation  $o$  of  $\mathbb{R}^n$  to  $o_{x_0}$ , then this implies that  $\sigma'(\vec{u})$  translates  $o$  to  $o_x$ . If  $\sigma'(\vec{u}_0)$  translates the negative  $-o$  of the standard orientation of  $\mathbb{R}^n$  to  $o_{x_0}$ , then for the modified chart

$$\tau = \sigma \circ J^{-1}: J(U) \rightarrow M, \quad J(u_1, u_2, \dots, u_n) = (-u_1, u_2, \dots, u_n),$$

$\tau'(J(\vec{u}_0))$  translates the standard orientation  $o$  to  $o_{x_0}$ . The same argument as before shows that  $\tau'(J(\vec{u}))$  translates  $o$  to  $o_x$ . We conclude that either  $\sigma$  or  $\sigma \circ J$  is an orientation compatible chart. This proves that every point is contained in an orientation compatible chart. Combining these charts together, we get an orientation compatible atlas. As argued before the proposition, the derivatives of the transition maps in this atlas have positive determinants.

Conversely, suppose there is an atlas  $\{\sigma_i\}$ , such that the derivatives of all the transition maps have positive determinants. For any  $x \in M$ , take a chart  $\sigma_i$  containing  $x$  and define the orientation  $o_x$  to be the translation of  $o$  via the isomorphism  $\sigma'_i(\vec{u}): \mathbb{R}^n \cong T_x M$ ,  $x = \sigma_i(\vec{u})$ . Since  $\det \varphi'_{ji} > 0$ , the transition maps preserve orientations, so that  $o_x$  is independent of the choice of chart. The construction of  $o_x$  also implies that the compatibility condition in Definition 14.5.1 is satisfied. Therefore we get an orientation for  $M$ .  $\square$

**Example 14.5.1.** The tangent space of  $S^1$  at a point is 1-dimensional, and its orientation is given by one basis vector. Therefore an orientation of  $S^1$  is given by a nonzero tangent field. An example is to choose the tangent vector at  $\vec{x} \in S^1$  to be the counterclockwise rotation  $R(\vec{x})$  of  $\vec{x}$  by 90 degrees.

To verify the compatibility condition, we cover the unit circle by two charts

$$\sigma_0(\theta) = (\cos \theta, \sin \theta): (0, 2\pi) \rightarrow S^1, \quad \sigma_1(\theta) = (\cos \theta, \sin \theta): (-\pi, \pi) \rightarrow S^1.$$

The derivative isomorphism of the first chart is

$$\mathbb{R} \xrightarrow[\cong]{\sigma'_0(\theta)} T_{\vec{x}}S^1: 1 \mapsto (-\sin \theta, \cos \theta) = R(\vec{x}), \quad \vec{x} = \sigma_0(\theta).$$

Therefore the derivative translates the standard orientation (given by the basis 1 of  $\mathbb{R}$ ) to the orientation of  $T_{\vec{x}}S^1$  (given by the basis  $R(\vec{x})$  of  $T_{\vec{x}}S^1$ ). The same happens to the chart  $\sigma_1$ . This verifies the compatibility, and also shows that the charts are orientation compatible.

We may also choose charts

$$\sigma_0(\theta) = (\cos \theta, \sin \theta): (0, 2\pi) \rightarrow S^1, \quad \sigma_1(\theta) = (\sin \theta, \cos \theta): (0, 2\pi) \rightarrow S^1.$$

The transition map has two parts

$$\varphi_{10}(\theta) = \begin{cases} \frac{\pi}{2} - \theta & : \left(0, \frac{\pi}{2}\right) \rightarrow \left(0, \frac{\pi}{2}\right), \\ \frac{5\pi}{2} - \theta & : \left(\frac{\pi}{2}, 2\pi\right) \rightarrow \left(\frac{\pi}{2}, 2\pi\right), \end{cases}$$

and  $\varphi'_{10} = -1 < 0$  on both parts. However, the negative sign does not mean the circle is not orientable. It only means that we need to compose the  $\sigma_1$  with  $J(\theta) = -\theta$  to get new  $\sigma_1$

$$\sigma_1(\theta) = (\sin(-\theta), \cos(-\theta)) = (-\sin \theta, \cos \theta): (-2\pi, 0) \rightarrow S^1.$$

Then the derivative of the transition between the two charts is +1.

Similar situation happened in Example 13.2.7, where we carefully choose the order of variables in the parameterizations, to make sure that all the pieces give the same normal vector, and the derivatives of the transition maps have positive determinants.

**Example 14.5.2.** Consider the open Möbius band in Example 14.1.4. The transition map has two parts, with  $\det(\varphi_{10}^+)' = 1$  and  $\det(\varphi_{10}^-)' = -1$ . Unlike Example 14.5.1, no adjustment of the orientation for the charts can make the determinant positive everywhere. Specifically, since the charts are connected, if we change the orientation for  $\sigma_1$ , then we will only get  $\det(\varphi_{10}^+)'$  negative and  $\det(\varphi_{10}^-)'$  positive. The problem is that any adjustment will change both signs simultaneously, and therefore cannot make both positive. The Möbius band is not orientable.

**Example 14.5.3.** The real projective space  $\mathbb{R}P^2$  in Example 14.1.5 is covered by three connected charts  $\sigma_i$ ,  $i = 0, 1, 2$ . The transition maps are

$$\varphi_{10}(u, v) = \left(\frac{1}{u}, \frac{v}{u}\right), \quad \varphi_{20}(u, v) = \left(\frac{1}{v}, \frac{u}{v}\right), \quad \varphi_{21}(u, v) = \left(\frac{u}{v}, \frac{1}{v}\right).$$

We have

$$\det \varphi'_{10} = -\frac{1}{u^3}, \quad \det \varphi'_{20} = \frac{1}{v^3}, \quad \det \varphi'_{21} = -\frac{1}{v^3}.$$

In fact, each transition map is broken into two parts, and  $\det \varphi'_{ji}$  have different signs on the two parts. Similar to Example 14.5.2, the real projective space  $\mathbb{R}P^2$  is not orientable.



Exercise 14.93. Show that the sphere  $S^n$  is orientable.

Exercise 14.94. Show that the real projective space  $\mathbb{R}P^n$  in Example 14.1.5 is orientable if and only if  $n$  is odd.

Exercise 14.95. Show that the complex projective space  $\mathbb{C}P^n$  in Example 14.1.6 is always orientable.

Exercise 14.96. Is the Klein bottle in Exercise 14.11 orientable?

Exercise 14.97. Prove that the special linear group  $SL(2)$  in Exercise 8.67 is orientable by finding an atlas so that the determinants of the derivatives of the transition maps are positive.

Exercise 14.98. Show that the mapping torus in Exercise 14.12 is orientable if and only if  $\det L > 0$ .

## Orientation of Manifold with Boundary

Definition 14.5.1 of orientation of manifold can be directly applied to manifold with boundary. All the discussions, including the local translation of orientation, the translation of orientation along curves, most of Proposition 14.5.2, and the remark after the proposition, are still valid. The only problem is that, at boundary points, the charts in Proposition 14.5.2 cannot be restricted to open subsets of the *upper* half space  $\mathbb{R}_+^n$ . For example, the atlas given in Example 14.1.7 for  $M = [0, 1]$  satisfies  $\det \varphi'_{21} = -1 < 0$ , while the interval should be orientable. However, if we insist that  $U_1$  and  $U_2$  must be of the form  $[0, a)$ , then it will necessarily follow that  $\varphi_{21}$  is a decreasing function and therefore has negative determinant.

The technical reason for the problem is that, in proving Proposition 14.5.2, we need to sometimes reverse the orientation of open subsets of  $\mathbb{R}^n$  by composing with  $J$ . Since  $J$  takes open subsets of  $\mathbb{R}_+^n$  to open subsets of  $\mathbb{R}_+^n$  only if  $n > 1$ , our problem occurs only for 1-dimensional manifold with boundary. So we have the following modified version of Proposition 14.5.2.

**Proposition 14.5.3.** *A manifold with boundary is orientable if and only if there is an atlas, in which the domain of each chart is an open subset of either  $\mathbb{R}^n$  or  $\mathbb{R}_+^n$ , such that the determinants of the derivatives of the transition maps are all positive. In case the manifold is 1-dimensional, we may also need to use open subsets of  $\mathbb{R}_- = (-\infty, 0]$  for the charts.*

In Green's Theorem, Stokes' Theorem and Gauss' Theorem in Sections 13.4, 13.5 and 13.6, the orientation of the boundary is defined by using the outward normal vector. In general, at a boundary point  $x_0$  of  $M$ , we may define an outward vector to be any vector (see Figure 14.3.1)

$$\vec{n}_{\text{out}} = [\gamma: (a, 0] \rightarrow M, \gamma(0) = x_0] \notin T_{x_0} \partial M.$$

Note that the manifold  $M$  may not carry an inner product, so that we cannot talk about outward *normal* vector in general. Consequently, the outward vector is not necessarily unique. Still, we may define the orientation of the boundary using the earlier idea.

**Definition 14.5.4.** An orientation of a manifold  $M$  with boundary induces an orientation on  $\partial M$ , by requiring that an ordered basis  $\{X_1, X_2, \dots, X_{n-1}\}$  of  $T_x \partial M$  gives the induced orientation if the ordered basis  $\{\vec{n}_{\text{out}}, X_1, X_2, \dots, X_{n-1}\}$  of  $T_x M$  gives the orientation of  $M$ .

There are two things we need to do to justify the definition. The first is the choices of the outward vector and the ordered basis  $\{X_1, X_2, \dots, X_{n-1}\}$ . By Proposition 14.3.4 and the subsequent discussion, for a chart  $\sigma: U \subset \mathbb{R}_+^n \rightarrow M$  around  $x_0 \in \partial M$ , we have an invertible map

$$(\sigma^{-1})'(x_0): (T_{x_0} M; T_{x_0+} M, T_{x_0-} M, T_{x_0} \partial M) \rightarrow (\mathbb{R}^n; \mathbb{R}_+^n, \mathbb{R}_-^n, \mathbb{R}^{n-1} \times 0). \quad (14.5.2)$$

The outward vector simply means that  $\vec{n}_{\text{out}} \in T_{x_0-} M - T_{x_0} \partial M$ . This is equivalent to that

$$(\sigma^{-1})'(x_0)(\vec{n}_{\text{out}}) \in \mathbb{R}_-^n - \mathbb{R}^{n-1} \times 0$$

has negative last coordinate  $(\vec{n}_{\text{out}})_n < 0$ . Since any two vectors  $\vec{w}_1, \vec{w}_2 \in \mathbb{R}_-^n - \mathbb{R}^{n-1} \times 0$  are related by

$$\vec{w}_1 = \lambda \vec{w}_2 + \vec{v}, \quad \lambda > 0, \quad \vec{v} \in \mathbb{R}^{n-1} \times 0,$$

we see that any two outward vectors  $\vec{n}_1, \vec{n}_2$  are related by

$$\vec{n}_1 = \lambda \vec{n}_2 + X, \quad \lambda > 0, \quad X \in T_{x_0} \partial M.$$

Then two choices  $\{\vec{n}_1, X_1, X_2, \dots, X_{n-1}\}$  and  $\{\vec{n}_2, Y_1, Y_2, \dots, Y_{n-1}\}$  are related by the matrix

$$\begin{pmatrix} \lambda & * & * & \cdots & * \\ 0 & a_{11} & a_{12} & \cdots & a_{1(n-1)} \\ 0 & a_{21} & a_{22} & \cdots & a_{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & a_{(n-1)1} & a_{(n-1)2} & \cdots & a_{(n-1)(n-1)} \end{pmatrix} = \begin{pmatrix} \lambda & * \\ 0 & A \end{pmatrix}.$$

Here  $A$  is the matrix between  $\{X_1, X_2, \dots, X_{n-1}\}$  and  $\{Y_1, Y_2, \dots, Y_{n-1}\}$ . By

$$\det \begin{pmatrix} \lambda & * \\ 0 & A \end{pmatrix} = \lambda \det A,$$

and  $\lambda > 0$ , we see that  $\{\vec{n}_1, X_1, X_2, \dots, X_{n-1}\}$  and  $\{\vec{n}_2, Y_1, Y_2, \dots, Y_{n-1}\}$  give the same orientation if and only if  $\{X_1, X_2, \dots, X_{n-1}\}$  and  $\{Y_1, Y_2, \dots, Y_{n-1}\}$  give the same orientation.

The second is to verify that the induced orientation on the boundary satisfies the compatibility condition in Definition 14.5.1. It is sufficient to show that  $\sigma'(\vec{u}) \circ$

$\sigma'(\vec{u}_0)^{-1}$  takes an outward vector at  $x_0$  to an outward vector at  $x$ . Since within one chart, we have the correspondence (14.5.2) of the pieces of the tangent space at every point in the chart, we have

$$\begin{aligned} T_{x_0}M - T_{x_0}\partial M &= \sigma'(\vec{u}_0)(\mathbb{R}^n - \mathbb{R}^{n-1} \times 0), \\ T_xM - T_x\partial M &= \sigma'(\vec{u})(\mathbb{R}^n - \mathbb{R}^{n-1} \times 0). \end{aligned}$$

Therefore the isomorphism  $\sigma'(\vec{u}) \circ \sigma'(\vec{u}_0)^{-1}$  preserves the outward vector.

Now we find orientation compatible charts for the boundary. Suppose  $\sigma: U \subset \mathbb{R}_+^n \rightarrow M$  is an orientation compatible chart around  $x_0 \in \partial M$ . Then the restriction

$$\tau = \sigma|_{\mathbb{R}^{n-1} \times 0}: V = U \cap \mathbb{R}^{n-1} \times 0 \rightarrow \partial M$$

is a chart of the boundary. Let  $\{X_1, X_2, \dots, X_{n-1}\}$  be an ordered basis of  $T_{x_0}\partial M$ . Then  $(\sigma^{-1})'(x_0)(X_i) = ((\tau^{-1})'(x_0)(X_i), 0) \in \mathbb{R}^{n-1} \times 0$ , and the requirement in Definition 14.5.4 is that  $(\sigma^{-1})'(x_0)$  translates  $\{\vec{n}_{\text{out}}, X_1, X_2, \dots, X_{n-1}\}$  to a positively oriented basis in  $\mathbb{R}^n$ . This means that

$$\det \begin{pmatrix} * & (\tau^{-1})'(X_1) & (\tau^{-1})'(X_2) & \cdots & (\tau^{-1})'(X_{n-1}) \\ (\vec{n}_{\text{out}})_n & 0 & 0 & \cdots & 0 \end{pmatrix} > 0.$$

Since  $(\vec{n}_{\text{out}})_n < 0$ , we get

$$(-1)^n \det((\tau^{-1})'(X_1) \ (\tau^{-1})'(X_2) \ \cdots \ (\tau^{-1})'(X_{n-1})) > 0.$$

This implies that the chart  $\tau$  is  $(-1)^n$ -compatibly oriented. In other words, if  $\{\sigma_i\}$  is an orientation compatible atlas of  $M$ , then for even  $n$ ,  $\{\sigma_i|_{\mathbb{R}^{n-1} \times 0}\}$  is an orientation compatible atlas of  $\partial M$ , and for odd  $n$ ,  $\{\sigma_i|_{\mathbb{R}^{n-1} \times 0} \circ J\}$  is an orientation compatible atlas of  $\partial M$ .

**Example 14.5.4.** Consider the modified atlas for  $M = [0, 1]$

$$\begin{aligned} \sigma_1(t) &= t: U_1 = [0, 1) \rightarrow M_1 = [0, 1) \subset M, \\ \sigma_2(t) &= 1 + t: U_2 = (-1, 0] \rightarrow M_2 = (0, 1] \subset M, \end{aligned}$$

that gives the orientation of  $M$ . The orientation on the 0-dimensional manifold  $\partial M = \{0, 1\}$  simply assigns  $\pm$  to the points. At  $0 \in \partial M$ , we note that  $n_{\text{out}} = -1$  for  $U_1$ , which is opposite to the orientation of  $[0, 1]$ . Therefore we correct this by assigning  $-$  to the point 0. At  $1 \in \partial M$ , we note that  $n_{\text{out}} = 1$  for  $U_2$ , which is the same as the orientation of  $[0, 1]$ . Therefore we keep this by assigning  $+$  to the point 1. The boundary  $\partial M = \{0, 1\}$  has the compatible orientation that assigns  $-$  to 0 and  $+$  to 1.

**Example 14.5.5.** Let  $f$  be a continuously differentiable function on an open subset  $U \subset \mathbb{R}^{n-1}$ . Then the part of  $\mathbb{R}^n$  over the graph of  $f$

$$M = \{(\vec{u}, z): z \geq f(\vec{u})\} \subset \mathbb{R}^n$$

is a manifold with boundary, and the boundary  $\partial M$  is the graph of  $f$ . The standard orientation of  $\mathbb{R}^n$  restricts to an orientation on  $M$ .

We have a natural diffeomorphism

$$\sigma(\vec{u}, t) = (\vec{u}, t + f(\vec{u})): U \times [0, +\infty) \rightarrow M,$$

that can be regarded as a chart covering the whole  $M$ . Since the derivative

$$\sigma'(\vec{u}_0, t_0)(\vec{v}, s) = (\vec{v}, s + f'(\vec{u}_0)(\vec{v})): T_{(\vec{u}_0, t_0)}(U \times [0, +\infty)) \cong \mathbb{R}^n \rightarrow T_{\vec{x}_0} M \cong \mathbb{R}^n$$

has determinant 1, the chart is compatible with the orientation of  $M$ . The restriction of the chart to the boundary

$$\tau(\vec{u}) = (\vec{u}, f(\vec{u})): U \rightarrow \partial M$$

is then  $(-1)^n$ -compatible with the induced orientation on the graph of  $f$ . This accounts for the sign  $(-1)^n$  in (13.6.2) for the integration along the bottom piece.

For the orientation used in the integration along the top piece used in (13.6.1), we need to consider the part of  $\mathbb{R}^n$  below the graph of  $f$

$$M = \{(\vec{u}, z): z \leq f(\vec{u})\} \subset \mathbb{R}^n.$$

The chart should be changed to

$$\sigma(\vec{u}, t) = (\vec{u}, -t + f(\vec{u})): U \times [0, +\infty) \rightarrow M,$$

and its derivative has determinant  $-1$ . Therefore restriction of the chart to the boundary

$$\tau(\vec{u}) = (\vec{u}, f(\vec{u})): U \rightarrow \partial M$$

is  $(-1)^{n-1}$ -compatible with the induced orientation on the graph of  $f$ . This matches the sign  $(-1)^{n-1}$  in (13.6.1).

**Exercise 14.99.** In Example 14.5.5, the last coordinate is a function of the first  $n - 1$  coordinates. If the last coordinate is changed to the  $i$ -th coordinate, how is the restriction of  $\sigma$  to  $\mathbb{R}^{n-1} \times 0$  compatible with the induced orientation on  $\partial M$ ?

**Exercise 14.100.** Suppose  $\sigma: U \rightarrow M$  is an orientation compatible chart around  $x_0 \in \partial M$ , where  $U$  is an open subset of the *lower* half space

$$\mathbb{R}_-^n = \{(x_1, x_2, \dots, x_{n-1}, x_n): x_n \leq 0\}.$$

How is the restriction of  $\sigma$  to  $\mathbb{R}^{n-1} \times 0$  compatible with the induced orientation on  $\partial M$ ? Then use such a chart to explain the sign  $(-1)^{n-1}$  in (13.6.1).

**Exercise 14.101.** The sphere  $S^n$  is the boundary of the unit ball  $B^{n+1}$ , which inherits the standard orientation of  $\mathbb{R}^{n+1}$ . Describe the induced orientation on the sphere.

**Exercise 14.102.** Suppose  $M$  and  $N$  are orientable manifolds with boundary. Prove that  $\partial(M \times N) = \partial M \times N \cup (-1)^m M \times \partial N$ ,  $m = \dim M$ .

## Chapter 15

# Field on Manifold

## 15.1 Tangent Field

In this chapter, we always assume that manifolds, functions, maps, and whatever quantities on manifolds are continuously differentiable of sufficiently high order.

A *tangent field* on a manifold  $M$  assigns a tangent vector  $X(x) \in T_x M$  to each point  $x \in M$ . For a function  $f$  on  $M$ ,  $X(f)$  is again function on  $M$ , with the value at  $x$  obtained by applying  $X(x)$  to the function  $f$

$$X(f)(x) = X(x)(f).$$

### Flow

A flow is a time dependent movement. Suppose the point  $x \in M$  is at its original location at time  $t = 0$  and moves to  $\xi(x, t)$  at time  $t$ . Then the flow is described by a map

$$\xi(x, t): M \times [0, +\infty) \rightarrow M, \quad \xi(x, 0) = x.$$

The condition  $\xi(x, 0) = x$  means the initial position of the point. The range  $[0, +\infty)$  for the time  $t$  may be too optimistic, because flows often cannot last forever. On the other hand, it is often convenient to allow negative time. For example, if we reset time  $t_0$  to be the initial time, then the flow is really defined for  $t \geq -t_0$ . In the subsequent discussion, we will assume that  $t \in (-\delta, \delta)$ , and  $\delta$  may depend on  $x$ .

For fixed  $t$ , the map  $\xi^t: M \rightarrow M$  defined by  $\xi^t(x) = \xi(x, t)$  can be considered as the movement from the original positions at time  $t$ . For fixed  $x$ , the curve  $\xi^x: (-\delta, \delta) \rightarrow M$  defined by  $\xi^x(t) = \xi(x, t)$  can be considered as the track of the movement of a point  $x$ . Either viewpoint is useful for understanding flows.

**Example 15.1.1.** The rightward movement of the plane at constant speed  $v$  is given by

$$\xi((x, y), t) = (x + vt, y): \mathbb{R}^2 \times [0, +\infty) \rightarrow \mathbb{R}^2.$$

The movement  $\xi^t$  at time  $t$  shifts the whole plane rightward by distance  $vt$ . The track  $\xi^{(x, y)}$  is the horizontal line of height  $y$ .

If we restrict the flow to the unit disk  $U = \{(x, y): x^2 + y^2 < 1\}$ , then  $\xi((x, y), t)$  is defined only when it lies inside  $U$ . This means that, for fixed  $(x, y) \in U$ ,  $\xi((x, y), t)$  is defined only for  $t \in \left(\frac{1}{v}(-\sqrt{1-y^2} + x), \frac{1}{v}(\sqrt{1-y^2} + x)\right)$ .

**Example 15.1.2.** The rotation of the plane around the origin at constant angular speed  $v$  is given by

$$\xi((x, y), t) = (x \cos vt - y \sin vt, x \sin vt + y \cos vt): \mathbb{R}^2 \times [0, +\infty) \rightarrow \mathbb{R}^2.$$

The movement  $\xi^t$  at time  $t$  rotates the whole plane by angle  $vt$ . The track  $\xi^{(x, y)}$  is a circle entered at the origin.

Our usual perception of the flow is the *velocity*  $\frac{\partial \xi}{\partial t}$ . Since the velocity is measured at the location  $\xi(x, t)$  of  $x$  at time  $t$ , not the original location  $x$  at  $t = 0$ ,

the corresponding velocity tangent field  $X$  satisfies

$$X(\xi(x, t), t) = \frac{\partial \xi}{\partial t}(x, t).$$

If the map  $\xi^t$  is invertible for every  $t$ , then the velocity tangent field at time  $t$  is

$$X^t(x) = X(x, t) = \frac{\partial \xi}{\partial t}((\xi^t)^{-1}(x), t).$$

Conversely, suppose  $X^t(x) = X(x, t)$  is a (perhaps time dependent) tangent field. Then we may recover the flow  $\xi(x, t)$  by solving

$$\frac{\partial \xi}{\partial t}(x, t) = X(\xi(x, t), t), \quad \xi(x, 0) = x. \quad (15.1.1)$$

For each fixed  $x$ , this is an initial value problem for an ordinary differential equation in  $t$ . The solution is the track  $\xi^x(t)$  of the movement of the point  $x$  in the flow. By the theory of ordinary differential equation, if the tangent field satisfies the Lipschitz condition (a consequence of continuous differentiability, for example), then the equation has unique solution for  $t \in (-\delta, \delta)$  for some small  $\delta$  that may depend on  $x$ .

**Definition 15.1.1.** Suppose  $X(x, t)$  is a (perhaps time dependent) tangent field on a manifold  $M$ . The *flow* induced by the tangent field is a map  $\xi(x, t) \in M$  defined for  $x \in M$  and small  $t$ , such that (15.1.1) holds. If  $\xi$  is defined for all  $t \in \mathbb{R}$ , then the flow is *complete*.

**Example 15.1.3.** The rotation of the plane around the origin in Example 15.1.2 has

$$\frac{\partial \xi}{\partial t}((x, y), t) = (-vx \sin vt - vy \cos vt, vx \cos vt - vy \sin vt).$$

Since this is the tangent field  $X$  at  $\xi((x, y), t)$ , or

$$\frac{\partial \xi}{\partial t}((x, y), t) = X(x \cos vt - y \sin vt, x \sin vt + y \cos vt),$$

we get

$$X(x, y) = (-vy, vx).$$

Conversely, given the tangent field  $X(x, y) = (-vy, vx)$ , the corresponding flow  $\xi((x, y), t) = (f((x, y), t), g((x, y), t))$  can be found by solving the initial value problem

$$\frac{\partial f}{\partial t} = -vg, \quad \frac{\partial g}{\partial t} = vf, \quad f = x \text{ and } g = y \text{ at } t = 0.$$

Eliminating  $g$ , we get

$$\frac{\partial^2 f}{\partial t^2} + v^2 f = 0.$$

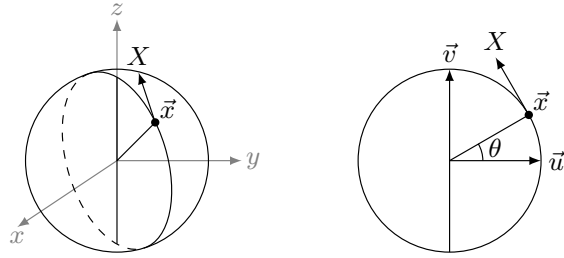
The solution is  $f = A \cos vt + B \sin vt$ . Substituting into the original equation, we get  $g = A \sin vt - B \cos vt$ . Then the initial values of  $f$  and  $g$  imply  $A = x, B = -y$ , so that  $f = x \cos vt - y \sin vt$  and  $g = x \sin vt + y \cos vt$ .

**Example 15.1.4.** On the unit sphere  $S^2$ , consider the tangent field at  $\vec{x} = (x, y, z) \in S^2$  that points “upward” and has length  $\sqrt{1 - z^2}$ . In Figure 15.1.1, we have

$$\vec{u} = \frac{(x, y, 0)}{\sqrt{1 - z^2}}, \quad \vec{v} = (0, 0, 1), \quad z = \sin \theta.$$

Therefore

$$X(x, y, z) = \sqrt{1 - z^2}(-\vec{u} \sin \theta + \vec{v} \cos \theta) = (-xz, -yz, 1 - z^2).$$



**Figure 15.1.1.** Flow from south to north.

The corresponding flow is expected to go vertically from the south pole  $(0, 0, -1)$  to the north pole  $(0, 0, 1)$ . The flow  $\xi((x, y, z), t) = (f, g, h)$  can be solved from

$$\frac{\partial f}{\partial t} = -fh, \quad \frac{\partial g}{\partial t} = -gh, \quad \frac{\partial h}{\partial t} = 1 - h^2, \quad (f, g, h)_{t=0} = (x, y, z).$$

We get

$$h = \frac{(z+1)e^{2t} + z - 1}{(z+1)e^{2t} - z + 1}, \quad f = xe^{-\int_0^t h}, \quad g = ye^{-\int_0^t h}.$$

A simpler description of the flow  $\xi$  is based on the observation that the flow is contained in the half great arc from the south pole to the north pole. Each half great arc is parameterised by  $\theta \in I = [-\frac{\pi}{2}, \frac{\pi}{2}]$ , and the flow in the half great arc may be translated to a flow  $\xi(\theta, t)$  in the interval  $I$ . The tangent field is translated to  $X(\theta) = \sqrt{1 - z^2} = \cos \theta$  on the interval, and the flow can be obtained by solving

$$\frac{\partial \xi}{\partial t} = \cos \xi, \quad \xi(\theta, 0) = \theta.$$

The solution  $\xi$  satisfies

$$\frac{\sin \xi + 1}{\sin \xi - 1} = \frac{\sin \theta + 1}{\sin \theta - 1} e^{2t},$$

from which we can get the formula for  $\xi$ .

**Exercise 15.1.** Describe the tangent field of the flow.

1.  $X$  is the rotation of  $S^2$  around the (north, south) axis at angular velocity  $v$ .
2.  $X$  moves any point  $(x_0, y_0)$  on  $\mathbb{R}^2$  along the curve  $y - y_0 = (x - x_0)^2$  such that the speed in the  $y$ -direction is 1.

**Exercise 15.2.** Find the flow for the tangent field.



1.  $X = \vec{a}$  is a constant vector field on  $\mathbb{R}^n$ .
2.  $X = \vec{a} + t\vec{b}$  on  $\mathbb{R}^n$ .
3.  $X = (x, -y)$  on  $\mathbb{R}^2$ .

**Exercise 15.3.** Let  $F: M \rightarrow N$  be a map. We say that (time independent) tangent fields  $X$  on  $M$  and  $Y$  on  $N$  are *F-related* if  $Y_{F(x)} = F_*X_x$  for all  $x \in M$ . Prove that the flows  $\xi^t$  and  $\eta^t$  of  $X$  and  $Y$  are also related by  $\eta^t(F(x)) = F(\xi^t(x))$ . In particular, this means that diffeomorphisms preserve flows of corresponding tangent fields.

## Flow Diffeomorphism

A flow moves a point  $x$  from its location  $x_1 = \xi(x, t_1)$  at time  $t_1$  to its location  $x_2 = \xi(x, t_2)$  at time  $t_2$ . This suggests a map

$$\xi^{t_1 \rightarrow t_2}(x_1) = x_2: M \rightarrow M$$

that describes the movement of the location at time  $t_1$  to the location at time  $t_2$ . The immediate problem with the map is whether it is well defined, because we do not know whether  $x_1$  can be equal to  $\xi(x, t_1)$  for two different  $x$ . So we alternatively define  $\xi^{t_1 \rightarrow t_2}$  by solving the initial value problem (the initial time is  $t_1$ )

$$\frac{d\gamma(t)}{dt} = X(\gamma(t), t), \quad \gamma(t_1) = x_1,$$

and then evaluating at  $t_2$  to get  $x_2 = \gamma(t_2)$ . The process does not make use of  $x$  and gives a well defined map  $\xi^{t_1 \rightarrow t_2}$ .

Although our intuition of the flow may implicitly require  $t_1 < t_2$ , the theory of ordinary differential equation allows  $t$  to be earlier than the initial time. This means that  $\xi^{t_1 \rightarrow t_2}$  is also defined for  $t_1 > t_2$ . Moreover, since our earlier flow  $\xi^t$  has initial time  $t = 0$ , we have  $\xi^t = \xi^{0 \rightarrow t}$ .

Let  $x_1, x_2, x_3$  be the location of  $x$  at time  $t_1, t_2, t_3$ . Then

$$\xi^{t_2 \rightarrow t_3}(\xi^{t_1 \rightarrow t_2}(x_1)) = x_3 = \xi^{t_1 \rightarrow t_3}(x_1).$$

This reflects the intuition that moving the locations of  $x$  from time  $t_1$  to  $t_3$  is the same as first moving from time  $t_1$  to  $t_2$  and then moving from time  $t_2$  to  $t_3$ . The equality means

$$\xi^{t_2 \rightarrow t_3} \circ \xi^{t_1 \rightarrow t_2} = \xi^{t_1 \rightarrow t_3}. \quad (15.1.2)$$

The equality can be rigorously proved by the uniqueness of the solutions of ordinary differential equations with initial conditions.

By the definition of  $\xi^{t_1 \rightarrow t_2}$ , the map  $\xi^{t_1 \rightarrow t_1}$  is the identity. By taking  $t_3 = t_1$  in (15.1.2), we find that the map  $\xi^{t_1 \rightarrow t_2}$  is invertible, with  $\xi^{t_2 \rightarrow t_1}$  as the inverse. Then we further get  $\xi^{t_1 \rightarrow t_2} = \xi^{0 \rightarrow t_2} \circ \xi^{t_1 \rightarrow 0} = \xi^{t_2} \circ (\xi^{t_1})^{-1}$ .

**Theorem 15.1.2.** *The flow of a tangent field is a diffeomorphism between locations of points at any two moments.*

The diffeomorphism in Examples 15.1.3 is the rotation of the plane to itself. The diffeomorphism in Example 15.1.4 moves the horizontal circles up to different levels. The tangent fields in the two examples are *time independent*, in the sense that  $X$  depends only on the location  $x$  and not on the time  $t$ . In this case, the initial value problem (15.1.1) becomes

$$\frac{\partial \xi}{\partial t}(x, t) = X(\xi(x, t)), \quad \xi(x, 0) = x.$$

This implies that  $\gamma(t) = \xi(x, t - t_1) = \xi^{t-t_1}(x)$  is the solution of the following initial value problem

$$\frac{d\gamma(t)}{dt} = X(\gamma(t)), \quad \gamma(t_1) = x.$$

Therefore we conclude that  $\xi^{t_1 \rightarrow t_2}(x) = \gamma(t_2) = \xi^{t_2-t_1}(x)$ , and the equality (15.1.2) becomes

$$\xi^{s+t} = \xi^s \circ \xi^t \quad \text{for time independent flow.}$$

Since  $\xi^0$  is the identity map, we also know that  $\xi^{-t}$  is the inverse of  $\xi^t$ .

## 15.2 Differential Form

The exterior algebra of the cotangent space  $T_{x_0}^*M$  is

$$\Lambda T_{x_0}^*M = \Lambda^0 T_{x_0}^*M \oplus \Lambda^1 T_{x_0}^*M \oplus \Lambda^2 T_{x_0}^*M \oplus \cdots \oplus \Lambda^n T_{x_0}^*M.$$

Since  $T_{x_0}^*M$  is the dual of  $T_{x_0}M$ , a vector (called *k-vector*) in  $\Lambda^k T_{x_0}^*M$  is a multilinear alternating function on tangent vectors. In particular, we have

$$df_1 \wedge \cdots \wedge df_k(X_1, \dots, X_k) = \det(df_i(X_j)) = \det(X_j(f_i)).$$

A map  $F: M \rightarrow N$  induces the pullback linear transform  $F^*: T_{F(x_0)}^*N \rightarrow T_{x_0}^*M$ , which further induces an algebra homomorphism  $F^*: \Lambda T_{F(x_0)}^*N \rightarrow \Lambda T_{x_0}^*M$ . In terms of multilinear alternating functions of tangent vectors, the pullback is

$$F^*\omega(X_1, \dots, X_k) = \omega(F_*X_1, \dots, F_*X_k), \quad X_i \in T_{x_0}M.$$

**Exercise 15.4.** Extend Exercise 14.52. Let  $\omega_t \in \Lambda^k T_{x_0}^*M$  be a curve of  $k$ -vectors at  $x_0$ . Prove that

$$\left. \frac{d}{dt} \right|_{t=0} \omega_t = \rho \iff \left. \frac{d}{dt} \right|_{t=0} \omega_t(X_1, \dots, X_k) = \rho(X_1, \dots, X_k) \text{ for all } X_i \in T_{x_0}M,$$

and

$$\int_a^b \omega_t dt = \rho \iff \int_a^b \omega_t(X_1, \dots, X_k) dt = \rho(X_1, \dots, X_k) \text{ for all } X_i \in T_{x_0}M.$$

**Exercise 15.5.** Extend Exercise 14.53. Let  $X_{1t}, \dots, X_{kt} \in T_{x_0}M$  be curves of tangent vectors at  $x_0$ , and let  $\omega_t \in \Lambda^k T_{x_0}^*M$  be a curve of  $k$ -vectors at  $x_0$ . Prove the Leibniz rule

$$\begin{aligned} \left. \frac{d}{dt} \right|_{t=0} (\omega_t(X_{1t}, \dots, X_{kt})) &= \left( \left. \frac{d}{dt} \right|_{t=0} \omega_t \right) (X_{10}, \dots, X_{k0}) \\ &\quad + \sum_{j=1}^k \omega_0 \left( X_{1t}, \dots, \left. \frac{dX_{jt}}{dt} \right|_{t=0}, \dots, X_{kt} \right) \Big|_{t=0}. \end{aligned}$$

**Exercise 15.6.** Let  $X_{1t}, \dots, X_{kt} \in T_{x_0}M$  be curves of tangent vectors at  $x_0$ . Prove that for any  $\omega \in \Lambda^k T_{x_0}^*M$ , we have

$$\int_a^b \omega(X_{1t}, \dots, X_{kt}) dt = \omega \left( \int_a^b X_{1t} \wedge \dots \wedge X_{kt} dt \right)$$

This partially extends the second part of Exercise 14.51.

## Differential Form

A *differential  $k$ -form* on  $M$  assigns a  $k$ -vector  $\omega(x) \in \Lambda^k T_x^*M$  to each  $x \in M$ . A 0-form is simply a function on the manifold. A 1-form is a *cotangent field*. For example, any function  $f$  gives a 1-form  $df$ . The collection of all  $k$ -forms on  $M$  is denoted  $\Omega^k M$ . By the vector space structure on  $\Lambda^k T_x^*M$ , we may take linear combinations of differential forms  $\omega, \rho \in \Omega^k M$  with functions  $f, g$  as coefficients

$$(f\omega + g\rho)(x) = f(x)\omega(x) + g(x)\rho(x).$$

We also have the *wedge product* of differential forms

$$(\omega \wedge \rho)(x) = \omega(x) \wedge \rho(x): \Omega^k M \times \Omega^l M \rightarrow \Omega^{k+l} M,$$

and the wedge product is *graded commutative*

$$\rho \wedge \omega = (-1)^{kl} \omega \wedge \rho \text{ for } \omega \in \Omega^k M \text{ and } \rho \in \Omega^l M.$$

Therefore  $\Omega M = \bigoplus_{k=0}^n \Omega^k M$  is a graded commutative algebra over the commutative ring  $C^r(M) = \Omega^0 M$  of functions. Since any  $k$ -form is locally a sum of  $gdf_1 \wedge \dots \wedge df_k$ , the algebra  $\Omega M$  is locally generated by functions  $f$  and their differentials  $df$ .

A map  $F: M \rightarrow N$  induces the pullback of differential forms

$$F^*: \Omega^k N \rightarrow \Omega^k M.$$

For  $f \in \Omega^0 N$ , we have  $F^*f = f \circ F$ . We also have  $F^*df = d(f \circ F) = dF^*f$  by (14.4.1). Since the pullback is an algebra homomorphism, we get

$$\begin{aligned} F^*(gdf_1 \wedge \dots \wedge df_k) &= (F^*g)(F^*df_1) \wedge \dots \wedge (F^*df_k) \\ &= (g \circ F) d(f_1 \circ F) \wedge \dots \wedge d(f_k \circ F). \end{aligned} \quad (15.2.1)$$

**Example 15.2.1.** For a 1-form  $\omega = f_1(\vec{x})dx_1 + \dots + f_n(\vec{x})dx_n$  on  $\mathbb{R}^n$  and a curve  $\gamma(t) = (x_1(t), \dots, x_n(t)): (a, b) \rightarrow \mathbb{R}^n$ , the pullback (or restriction) of the 1-form to the curve is

$$\gamma^*\omega = f_1(\gamma(t))dx_1(t) + \dots + f_n(\gamma(t))dx_n(t) = [f_1(\gamma(t))x'_1(t) + \dots + f_n(\gamma(t))x'_n(t)]dt.$$

**Example 15.2.2.** Let  $\omega = f(x, y, z)dy \wedge dz + g(x, y, z)dz \wedge dx + h(x, y, z)dx \wedge dy$  be a 2-form on  $\mathbb{R}^3$ . Let  $\sigma(u, v) = (x(u, v), y(u, v), z(u, v)): U \rightarrow S \subset \mathbb{R}^3$  be a parameterisation of a surface  $S$  in  $\mathbb{R}^3$ . The pullback (or restriction) of  $f(x, y, z)dy \wedge dz$  to the surface is

$$\begin{aligned}\sigma^*(f(x, y, z)dy \wedge dz) &= f(\sigma(u, v))dy(u, v) \wedge dz(u, v) \\ &= f(\sigma(u, v))(y_u du + y_v dv) \wedge (z_u du + z_v dv) \\ &= f(\sigma(u, v))(y_u z_v du \wedge dv + y_v z_u dv \wedge du) \\ &= f(\sigma(u, v)) \frac{\partial(y, z)}{\partial(u, v)} du \wedge dv.\end{aligned}$$

The pullback of the 2-form to the surface is

$$\sigma^*\omega = \left( f(\sigma(u, v)) \frac{\partial(y, z)}{\partial(u, v)} + g(\sigma(u, v)) \frac{\partial(z, x)}{\partial(u, v)} + h(\sigma(u, v)) \frac{\partial(x, y)}{\partial(u, v)} \right) du \wedge dv.$$

This is the same as the calculation leading to the integral (13.2.4) of 2-form on a surface in  $\mathbb{R}^3$ .

**Example 15.2.3.** Let  $\sigma: U \rightarrow M$  and  $\tau: V \rightarrow M$  be two charts of an  $n$ -dimensional manifold. An  $n$ -form  $\omega$  has local expressions in terms of the coordinates in the charts

$$\omega = g du_1 \wedge \cdots \wedge du_n = h dv_1 \wedge \cdots \wedge dv_n,$$

where  $g$  and  $h$  are functions. If we regard  $g$  as a function on  $U$ , then  $g du_1 \wedge \cdots \wedge du_n$  is actually the pullback  $\sigma^*\omega$ . On the other hand, we can also consider  $g$  and  $u_1, \dots, u_n$  as functions on  $\sigma(U) \subset M$ , so that  $g du_1 \wedge \cdots \wedge du_n$  is the differential form  $\omega$  on  $M$ .

The transition map  $\varphi = \tau^{-1} \circ \sigma$  is simply coordinates of  $V$  expressed as functions of coordinates of  $U$ :  $v_i = v_i(u_1, \dots, u_n)$ . Then by (14.3.7), we have

$$dv_i = \frac{\partial v_i}{\partial u_1} + \cdots + \frac{\partial v_i}{\partial u_n}.$$

Then by (7.3.3) (also see Exercise 7.56), we get

$$dv_1 \wedge \cdots \wedge dv_n = (\det \varphi') du_1 \wedge \cdots \wedge du_n, \quad \varphi' = \frac{\partial(v_1, \dots, v_n)}{\partial(u_1, \dots, u_n)}.$$

This implies  $g = h \det \varphi'$ .

**Exercise 15.7.** The spherical coordinate gives a map

$$F(r, \theta, \phi) = (r \cos \phi \cos \theta, r \cos \phi \sin \theta, r \sin \phi).$$

1. Compute the pullback of  $dx$ ,  $dx \wedge dy$  and  $z^2 dx \wedge dy \wedge dz$ .
2. Find the differential form in  $x, y, z$  that pulls back to  $d\theta$ .

**Exercise 15.8.** Compute the pullback of a 2-form  $\sum_{i < j} f_{ij} dx_i \wedge dx_j$  to a surface in  $\mathbb{R}^n$ . The result should be related to the integral (13.2.6) of 2-form on a surface in  $\mathbb{R}^n$ . Then find the pullback of the differential form corresponding to the integral (13.3.2) of  $k$ -form on a  $k$ -dimensional submanifold in  $\mathbb{R}^n$ .

**Exercise 15.9.** Extend Example 15.2.3 to the relation between the expressions of a differential  $k$ -form on two charts.

## Exterior Derivative

The exterior derivative is a derivative operation on differential forms that extends the differential of functions.

**Definition 15.2.1.** The *exterior derivative* is a map  $d: \Omega^k M \rightarrow \Omega^{k+1} M$  satisfying the following properties.

1.  $d$  is a derivation on the graded algebra  $\Omega M$ .
2.  $df = [f] \in T^*M$  is the usual differential of function, and  $d(df) = 0$ .

Recall that  $\Omega M = \bigoplus_{k=0}^n \Omega^k M$  is a graded algebra. In general, a graded algebra  $A = \bigoplus A_k$  over  $\mathbb{R}$  is defined as an algebra, such that the linear combination (with  $\mathbb{R}$ -coefficient) of elements in  $A_k$  lies in  $A_k$ , and the product of elements in  $A_k$  and  $A_l$  lies in  $A_{k+l}$ . The first condition in Definition 15.2.1 means the following.

**Definition 15.2.2.** A *derivation* of degree  $p$  on a graded algebra  $A = \bigoplus A_k$  over  $\mathbb{R}$  is a linear map  $D: A \rightarrow A$ , such that  $D$  maps  $A_k$  to  $A_{k+p}$  and satisfies the *Leibniz rule*

$$D(ab) = (Da)b + (-1)^{pk}a(Db), \quad a \in A_k, b \in A_l.$$

Although  $\Omega M$  is actually a graded commutative algebra over  $C^r(M)$ , we ignore the graded commutativity and the function coefficient when we say that the exterior derivative is a derivation. The Leibniz rule means

$$d(\omega \wedge \rho) = d\omega \wedge \rho + (-1)^k \omega \wedge d\rho \text{ for } \omega \in \Omega^k M \text{ and } \rho \in \Omega^l M.$$

By  $d(df) = 0$ , the Leibniz rule implies

$$d(gdf_1 \wedge \cdots \wedge df_k) = dg \wedge df_1 \wedge \cdots \wedge df_k. \quad (15.2.2)$$

This shows that it is uniquely determined by the two properties in Definition 15.2.1. In fact, we have the following general result.

**Proposition 15.2.3.** Suppose a graded algebra  $A$  is generated by homogeneous elements  $a_i \in A_{k_i}$ . If  $D_1$  and  $D_2$  are two derivations on  $A$  satisfying  $D_1(a_i) = D_2(a_i)$  for all  $a_i$ , then  $D_1 = D_2$  on the whole  $A$ .

The generators mean that any element of  $A$  is a linear combination of products of generators. The differential forms  $\Omega M$  is generated by functions  $f$  and differentials  $df$  of functions. The Leibniz rule reduces the calculation of the derivation of the element to the derivations of the generators. The proposition then follows.

**Proposition 15.2.4.** For any  $F: M \rightarrow N$  and  $\omega \in \Omega N$ , we have  $F^*d\omega = dF^*\omega$ .

The proposition extends (14.4.1). For  $\omega = gdf_1 \wedge \cdots \wedge df_k$ , the proposition is

obtained by combining (15.2.1) and (15.2.2)

$$F^*d\omega = dF^*\omega = d(g \circ F) \wedge d(f_1 \circ F) \wedge \cdots \wedge d(f_k \circ F).$$

**Proposition 15.2.5.** *For any  $\omega \in \Omega M$ , we have  $d^2\omega = 0$ .*

This extends the property  $d(df) = 0$  in the definition. For  $\omega = gdf_1 \wedge \cdots \wedge df_k$ , the proposition is easily obtained by applying the formula (15.2.2) twice.

**Exercise 15.10.** Prove that the exterior derivative is a *local operator*: If  $\omega = \rho$  near  $x$ , then  $d\omega = d\rho$  near  $x$ .

**Exercise 15.11.** Prove that the linear combination of derivations of the same degree is still a derivation. In fact, we can use the ring coefficients in the linear combinations. This means that if  $D$  is a derivation on  $\Omega M$  and  $f$  is a function, then  $fD$  is also a derivation.

**Exercise 15.12.** Let  $D: A \rightarrow A$  be a linear map of degree  $p$ . Explain that to verify the Leibniz rule, it is sufficient to verify for the case that  $a$  and  $b$  are products of generators. In fact, it is sufficient to verify for the case that  $a$  is a generator and  $b$  is a product of generators.

Although we calculated the exterior derivative and showed its properties, we have yet to establish its existence.

Any chart  $\sigma: U \rightarrow M$  gives a standard basis of  $\Omega^k M$ , and we may express any  $k$ -form in terms of the standard basis in the unique way

$$\omega = \sum a_{i_1 \dots i_k} du_{i_1} \wedge \cdots \wedge du_{i_k},$$

with functions  $a_{i_1 \dots i_k}$  on  $\sigma(U)$  as coefficients. Then we use the calculation (15.2.2) to define

$$d\omega = \sum da_{i_1 \dots i_k} \wedge du_{i_1} \wedge \cdots \wedge du_{i_k} \quad (15.2.3)$$

Here the operator  $d$  on the left is what we try to establish, and  $da_{i_1 \dots i_k} = [a_{i_1 \dots i_k}]$  and  $du_{i_p} = [u_{i_p}]$  are the usual differentials of functions. For  $k = 0$ , the definition simply means  $df$  on the left is the usual differential. The following verifies  $d(df) = 0$

$$\begin{aligned} d(df) &= d\left(\frac{\partial f}{\partial u_1} du_1 + \cdots + \frac{\partial f}{\partial u_n} du_n\right) \\ &= d\left(\frac{\partial f}{\partial u_1}\right) \wedge du_1 + \cdots + d\left(\frac{\partial f}{\partial u_n}\right) \wedge du_n \\ &= \left(\sum_i \frac{\partial^2 f}{\partial u_i \partial u_1} du_i\right) \wedge du_1 + \cdots + \left(\sum_i \frac{\partial^2 f}{\partial u_i \partial u_n} du_i\right) \wedge du_n \\ &= \sum_{i < j} \left(\frac{\partial^2 f}{\partial u_i \partial u_j} - \frac{\partial^2 f}{\partial u_j \partial u_i}\right) du_i \wedge du_j = 0. \end{aligned}$$

We note that the formula (15.2.3) remains valid even if  $i_1, \dots, i_k$  is not in ascending order, or has repeated index. Then for  $\omega = f du_{i_1} \wedge \dots \wedge du_{i_k}$  and  $\rho = g du_{j_1} \wedge \dots \wedge du_{j_l}$ , the following verifies the Leibniz rule

$$\begin{aligned}
 d(\omega \wedge \rho) &= d(fg du_{i_1} \wedge \dots \wedge du_{i_k} \wedge du_{j_1} \wedge \dots \wedge du_{j_l}) \\
 &= d(fg) \wedge du_{i_1} \wedge \dots \wedge du_{i_k} \wedge du_{j_1} \wedge \dots \wedge du_{j_l} \\
 &= (gdf + fdg) \wedge du_{i_1} \wedge \dots \wedge du_{i_k} \wedge du_{j_1} \wedge \dots \wedge du_{j_l} \\
 &= df \wedge du_{i_1} \wedge \dots \wedge du_{i_k} \wedge g du_{j_1} \wedge \dots \wedge du_{j_l} \\
 &\quad + (-1)^k f du_{i_1} \wedge \dots \wedge du_{i_k} \wedge dg \wedge du_{j_1} \wedge \dots \wedge du_{j_l} \\
 &= d\omega \wedge \rho + (-1)^k \omega \wedge d\rho.
 \end{aligned}$$

This completes the verification that the operation  $d\omega$  defined by the local formula (15.2.3) satisfies the two properties in Definition 15.2.1 and therefore gives an exterior derivative on  $\sigma(U)$ .

For the global existence of the exterior derivative, it remains to explain that the formula (15.2.3) is independent of the choice of charts. Let  $d_\sigma$  be the exterior derivative on  $\sigma(U)$  given by (15.2.3). Let  $d_\tau$  be the exterior derivative on another chart  $\tau(V)$  given by a formula similar to (15.2.3). Then both  $d_\sigma$  and  $d_\tau$  are derivations on  $\Omega(\sigma(U) \cap \tau(V))$ , such that  $d_\sigma f = d_\tau f$  is the usual differential  $df$  of functions, and  $d_\sigma(df) = d_\tau(df) = 0$ . By Proposition 15.2.3, therefore, the two exterior derivatives are equal on the overlapping.

## Exterior Derivative on Euclidean Space

We identify the exterior derivative on Euclidean space by using the identification

$$\Omega^k \mathbb{R}^n = \oplus_{1 \leq i_1 < \dots < i_k \leq n} C^r(\mathbb{R}^n) dx_{i_1} \wedge \dots \wedge dx_{i_k} \cong C^r(\mathbb{R}^n)^{\binom{n}{k}}.$$

The first differential  $C^r(\mathbb{R}^n) \cong \Omega^0 \mathbb{R}^n \xrightarrow{d} \Omega^1 \mathbb{R}^n \cong C^r(\mathbb{R}^n)^n$  is given by

$$df = \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_n} dx_n \leftrightarrow \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right) = \nabla f.$$

This is the gradient of function. The last differential  $C^r(\mathbb{R}^n)^n \cong \Omega^{n-1} \mathbb{R}^n \xrightarrow{d} \Omega^n \mathbb{R}^n \cong C^r(\mathbb{R}^n)$  is given by

$$\begin{aligned}
 (f_1, \dots, f_n) \leftrightarrow \omega &= \sum f_i dx_1 \wedge \dots \wedge \widehat{dx_i} \wedge \dots \wedge dx_n \\
 \mapsto d\omega &= \left( \frac{\partial f_1}{\partial x_1} - \frac{\partial f_2}{\partial x_2} + \dots + (-1)^{n-1} \frac{\partial f_n}{\partial x_n} \right) dx_1 \wedge \dots \wedge dx_n \\
 \leftrightarrow &\frac{\partial f_1}{\partial x_1} - \frac{\partial f_2}{\partial x_2} + \dots + (-1)^{n-1} \frac{\partial f_n}{\partial x_n}.
 \end{aligned}$$

To identify this with more familiar operation, we make use of the Hodge dual operator and consider  $\Omega^1\mathbb{R}^n \xrightarrow{\star} \Omega^{n-1}\mathbb{R}^n \xrightarrow{d} \Omega^n\mathbb{R}^n$

$$\begin{aligned} (f_1, \dots, f_n) &\leftrightarrow \omega = f_1 dx_1 + \dots + f_n dx_n \\ &\mapsto \omega^\star = \sum (-1)^{i-1} f_i dx_1 \wedge \dots \wedge \widehat{dx_i} \wedge \dots \wedge dx_n \\ &\mapsto d(\omega^\star) = \left( \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \dots + \frac{\partial f_n}{\partial x_n} \right) dx_1 \wedge \dots \wedge dx_n \\ &\leftrightarrow \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \dots + \frac{\partial f_n}{\partial x_n} = \nabla \cdot (f_1, \dots, f_n). \end{aligned}$$

This is the divergence of field.

On  $\mathbb{R}^1$ , both the gradient and the divergence are the same derivative operation  $df(x) = f'(x)dx: \Omega^0\mathbb{R} \rightarrow \Omega^1\mathbb{R}$ .

On  $\mathbb{R}^2$ , the differential  $df = f_x dx + f_y dy \leftrightarrow (f_x, f_y): \Omega^0\mathbb{R}^2 \rightarrow \Omega^1\mathbb{R}^2$  is naturally identified with the gradient. The differential  $d(fdx + gdy) = (-f_y + g_x)dx \wedge dy: \Omega^1\mathbb{R}^2 \rightarrow \Omega^2\mathbb{R}^2$  (note that  $f$  is  $f_2$  and  $g$  is  $f_1$ ) is naturally identified with the divergence only after the transformation  $fdx + gdy \mapsto fdy - gdx$  by the Hodge dual operator.

On  $\mathbb{R}^3$ , there are two differentials

$$\Omega^0\mathbb{R}^3 \xrightarrow{d} \Omega^1\mathbb{R}^3 \xrightarrow{d} \Omega^2\mathbb{R}^3 \xrightarrow{d} \Omega^3\mathbb{R}^3.$$

The first one is the gradient, and the third one is the divergence up to the Hodge dual operation. So we use the Hodge dual operation to replace  $\Omega^2\mathbb{R}^3$  by  $\Omega^1\mathbb{R}^3$

$$\Omega^0\mathbb{R}^3 \xrightarrow[\text{grad}]{d} \Omega^1\mathbb{R}^3 \xrightarrow[\text{div}]{\star^{-1}d} \Omega^1\mathbb{R}^3 \xrightarrow[\text{div}]{d\star} \Omega^3\mathbb{R}^3.$$

In general, we have  $\star^{-1} = \pm\star$ , and the sign is always positive for odd dimensional vector space. Therefore  $\star^{-1}d = \star d$ , and has the following explicit formula

$$\begin{aligned} (f, g, h) &\leftrightarrow \omega = fdx + gdy + hdz \\ &\mapsto d\omega = (h_y - g_z)dy \wedge dz + (h_x - f_z)dx \wedge dz + (g_x - f_y)dx \wedge dy \\ &\mapsto (d\omega)^\star = (h_y - g_z)dx + (f_z - h_x)dy + (g_x - f_y)dz \\ &\leftrightarrow (h_y - g_z, f_z - h_x, g_x - f_y) = \nabla \times (f, g, h). \end{aligned}$$

This is the curl of field.

**Exercise 15.13.** Compute  $\omega \wedge \omega$ ,  $d\omega$ ,  $d\theta$ ,  $\omega \wedge \theta$ ,  $d\omega \wedge \omega$ ,  $\omega \wedge d\theta$ , and  $d(\omega \wedge \theta)$ .

1.  $\omega = yzdx + x^2dz$ ,  $\theta = z^2dx + x^2dy$ .
2.  $\omega = xydx + (x+y)dy$ ,  $\theta = dx \wedge dy + yz^2dx \wedge dz + xz^2dy \wedge dz$ .

**Exercise 15.14.** What is the meaning of  $d^2 = 0$  in terms of the gradient, the curl, and the divergence?

**Exercise 15.15.** Identify the exterior derivatives on  $\mathbb{R}^4$ .

**Exercise 15.16.** Find the explicit formula for the *Laplacian operator*  $\Delta f = (d(df)^\star)^\star$ .



## 15.3 Lie Derivative

Let  $X$  a time independent tangent field on a manifold  $M$ . Let  $\xi$  be the flow associated to  $X$ . By moving the points of  $M$ , the flow also moves anything on the manifold. For example, a function  $f(x)$  on  $M$  is moved to a new function  $\xi^{t*}f(x) = f(\xi^t(x))$ , and the move satisfies  $\xi^{s*}(\xi^{t*}f) = \xi^{s+t*}f$ . The change of function caused by the flow can be measured by the derivative

$$\left. \frac{d}{dt} \right|_{t=0} \xi^{t*}f(x) = \left. \frac{d}{dt} \right|_{t=0} f(\xi^t(x)) = X(f)(x).$$

In the second equality, we used the fact that for fixed  $x$ , the equivalence class of the curve  $\xi^t(x)$  is exactly the tangent vector  $X(x)$ . What we get is the *Lie derivative* of a function  $f$  along a tangent field  $X$

$$L_X f = X(f). \quad (15.3.1)$$

### Lie Derivative of Tangent Field

Let  $Y$  be another (time independent) tangent field. The flow  $\xi$  moves  $Y$  by applying the pushforward<sup>36</sup>  $\xi_*^{-t}: T_{\xi^t(x)}M \rightarrow T_x M$ . Here we keep track of the locations of tangent vectors, and a tangent vector at  $x$  is obtained by using  $\xi^{-t}$  to push a tangent vector at  $\xi^t(x)$ . This means

$$\xi_*^{-t}Y(x) = \xi_*^{-t}(Y(\xi^t(x))) \in T_x M. \quad (15.3.2)$$

For fixed  $x$ ,  $\xi_*^{-t}Y(x)$  is a curve in the vector space  $T_x M$ , and the change of tangent field  $Y$  caused by the flow can be measured by the derivative of this curve. This is the *Lie derivative* of  $Y$  along  $X$

$$L_X Y = \left. \frac{d}{dt} \right|_{t=0} \xi_*^{-t}Y(x) = \lim_{t \rightarrow 0} \frac{\xi_*^{-t}Y(x) - Y(x)}{t} \in T_x M.$$

To compute the Lie derivative, we apply the tangent vector curve (15.3.2) to a function  $f$  at  $x$

$$\begin{aligned} L_X Y(f) &= \left. \frac{d}{dt} \right|_{t=0} \xi_*^{-t}Y(x)(f) && \text{(by Exercise 14.51)} \\ &= \left. \frac{d}{dt} \right|_{t=0} \xi_*^{-t}(Y(\xi^t(x)))(f) && \text{(by (15.3.2))} \\ &= \left. \frac{d}{dt} \right|_{t=0} Y(\xi^t(x))(f \circ \xi^{-t}). && \text{(by (14.4.2))} \end{aligned}$$

We have  $L_X Y(f) = \partial_t \phi(0, 0) - \partial_s \phi(0, 0)$  for

$$\phi(t, s) = Y(\xi^t(x))(f \circ \xi^s).$$

<sup>36</sup>We use  $\xi^t$  to move contravariant quantities such as functions and cotangent vectors, and use  $\xi^{-t}$  to move covariant quantities such as tangent vectors. The reason will become clear in (15.3.5).

Since  $\phi(t, 0) = Y(\xi^t(x))(f) = Y(f)(\xi^t(x))$  is the value of the function  $Y(f)$  at  $\xi^t(x)$ , and for fixed  $x$ , the equivalence class of the curve  $\xi^t(x)$  is exactly the tangent vector  $X(x)$ , we have

$$\partial_t \phi(0, 0) = \left. \frac{d}{dt} \right|_{t=0} Y(f)(\xi^t(x)) = X(x)(Y(f)) = X(Y(f))(x).$$

On the other hand, since  $\phi(0, s) = Y(x)(f \circ \xi^s)$ , and the equivalence class of the curve  $\xi^s$  is exactly the tangent vector  $X$ , we may apply Exercise 14.52 to get

$$\partial_s \phi(0, 0) = Y(x) \left( \left. \frac{d}{ds} \right|_{s=0} f \circ \xi^s \right) = Y(x)(X(f)) = Y(X(f))(x).$$

Therefore

$$L_X Y(f) = X(Y(f)) - Y(X(f)).$$

The Lie derivative is the *Lie bracket* of two tangent fields

$$L_X Y = [X, Y] = XY - YX.$$

This is supposed to be the difference of two second order derivatives. It turns out that the second order parts cancel and only the first order derivative remains. See Exercise 15.17.

**Exercise 15.17.** Suppose  $X = a_1 \partial_{u_1} + \cdots + a_n \partial_{u_n}$  and  $Y = b_1 \partial_{u_1} + \cdots + b_n \partial_{u_n}$  in a chart. Prove that

$$[X, Y] = \sum_i \left[ \sum_j \left( a_j \frac{\partial b_i}{\partial u_j} - b_j \frac{\partial a_i}{\partial u_j} \right) \right] \partial_{u_i} = \sum_i [X(b_i) - Y(a_i)] \partial_{u_i}.$$

**Exercise 15.18.** Suppose  $F: M \rightarrow N$  is a map. Suppose tangent fields  $X_1$  and  $X_2$  are  $F$ -related (see Exercise 15.3), and  $Y_1$  and  $Y_2$  are also  $F$ -related. Prove that  $[X_1, Y_1]$  and  $[X_2, Y_2]$  are also  $F$ -related.

**Exercise 15.19.** Prove the equality  $[fX, gY] = fg[X, Y] + fX(g)Y - gY(f)X$ .

**Exercise 15.20.** Prove the *Jacobi identity*

$$[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0.$$

**Exercise 15.21.** Explain that the Lie derivative satisfies the Leibniz rule:

$$\begin{aligned} L_X(fg) &= (L_X f)g + f(L_X g), \\ L_X(fY) &= (L_X f)Y + fL_X Y, \\ L_X[Y, Z] &= [L_X Y, Z] + [Y, L_X Z]. \end{aligned}$$

The last equality is the Jacobi identity in Exercise 15.20.

**Exercise 15.22.** Explain that the Lie derivative satisfies  $L_X L_Y - L_Y L_X = L_{[X, Y]}$ , when applied to functions and tangent fields. The application to tangent fields is the Jacobi identity in Exercise 15.20.

**Exercise 15.23.** Suppose  $\xi$  and  $\eta$  are the flows associated to vector fields  $X$  and  $Y$ . Prove that the two flows commute  $\xi^t \circ \eta^s = \eta^s \circ \xi^t$  if and only if  $[X, Y] = 0$ .

**Exercise 15.24.** For vector fields  $X_1, \dots, X_k$  on  $M$ , prove that the following are equivalent.

1.  $[X_i, X_j] = 0$  for all  $i, j$ .
2. Near every point, there is a chart, such that  $X_i = \partial_{u_i}$ .

**Exercise 15.25.**

$$(\eta^{-s} \circ \xi^{-t} \circ \eta^s \circ \xi^t)(x) = x - ts[X, Y] + \dots$$

**Exercise 15.26.** Suppose  $\xi$  and  $\eta$  are the flows associated to vector fields  $X$  and  $Y$ . Prove that the equivalence class of  $\gamma(t) = \eta^{-\sqrt{t}} \circ \xi^{-\sqrt{t}} \circ \eta^{\sqrt{t}} \circ \xi^{\sqrt{t}}$  is the tangent vector  $[X, Y]$ .

## Lie Derivative of Differential Form

The flow  $\xi$  moves  $\omega \in \Omega^k M$  through the pullback  $\xi^{t*} : T_{\xi^t(x)}^* M \rightarrow T_x^* M$

$$\xi^{t*} \omega(x) = \xi^{t*}(\omega(\xi^t(x))) \in \Lambda^k T_x^* M. \quad (15.3.3)$$

The change of  $\omega$  caused by the flow is the *Lie derivative* of  $\omega$  along  $X$

$$L_X \omega(x) = \left. \frac{d}{dt} \right|_{t=0} \xi^{t*} \omega(x) = \lim_{t \rightarrow 0} \frac{\xi^{t*} \omega(x) - \omega(x)}{t} \in \Lambda^k T_x^* M.$$

For the special case  $\omega = df$  is the differential of a function, by the last part of Exercise 14.52 and (15.3.1), we have

$$L_X df(x) = \left. \frac{d}{dt} \right|_{t=0} d_x(\xi^{t*} f) = d_x \left( \left. \frac{d}{dt} \right|_{t=0} \xi^{t*} f \right) = d_x(L_X f) = d_x X(f).$$

This means

$$L_X df = dL_X f. \quad (15.3.4)$$

Since the pullback is an algebra homomorphism on the exterior algebra, it is easy to see that

$$\xi^{t*}(\omega \wedge \rho) = \xi^{t*} \omega \wedge \xi^{t*} \rho.$$

Then by the Leibniz rule, we get

$$L_X(\omega \wedge \rho) = L_X \omega \wedge \rho + \omega \wedge L_X \rho.$$

This shows that  $L_X$  is a derivation on  $\Omega M$  of degree 0. Then (15.3.1) and (15.3.4) uniquely determines the Lie derivative of differential forms

$$L_X(gdf_1 \wedge \dots \wedge df_k) = X(g)df_1 \wedge \dots \wedge df_k + \sum_{i=1}^n gdf_1 \wedge \dots \wedge dX(f_i) \wedge \dots \wedge df_k.$$

We may also get the Lie derivative  $L_X\omega$  as a multilinear alternating function on tangent vectors. We take  $k$  tangent fields  $Y_i$  and consider the function on  $M$

$$f = \omega(Y_1, \dots, Y_k) = \langle Y_1 \wedge \dots \wedge Y_k, \omega \rangle.$$

We move the function by the flow

$$\begin{aligned} \xi^{t*} f(x) &= f(\xi^t(x)) = \langle Y_1(\xi^t(x)) \wedge \dots \wedge Y_k(\xi^t(x)), \omega(\xi^t(x)) \rangle \\ &= \langle \xi_*^{-t}(Y_1(\xi^t(x))) \wedge \dots \wedge \xi_*^{-t}(Y_k(\xi^t(x))), \xi^{t*}(\omega(\xi^t(x))) \rangle \\ &= \langle \xi_*^{-t}Y_1(x) \wedge \dots \wedge \xi_*^{-t}Y_k(x), \xi^{t*}\omega(x) \rangle. \end{aligned} \quad (15.3.5)$$

Here the second equality follows from  $\langle F_*^{-1}X, F^*\rho \rangle = \langle F_*(F_*^{-1}X), \rho \rangle = \langle X, \rho \rangle$  for any tangent vector  $X$  and cotangent vector  $\rho$ . Moreover, the third equality follows from (15.3.2) and (15.3.3). Taking the derivative in  $t$  at  $t = 0$  on both sides, we get  $L_X f = X(f)$  on the left. Since the right side is multilinear in  $\xi_*^{-t}Y_i$  and  $\xi^{t*}\omega$ , we have the Leibniz rule (see Exercise 8.31)

$$X(f) = L_X f = \sum_{i=1}^k \langle Y_1 \wedge \dots \wedge L_X Y_i \wedge \dots \wedge Y_k, \omega \rangle + \langle Y_1 \wedge \dots \wedge Y_k, L_X \omega \rangle.$$

By  $L_X Y_i = [X, Y_i]$ , this implies

$$L_X \omega(Y_1, \dots, Y_k) = X(\omega(Y_1, \dots, Y_k)) - \sum_{i=1}^k \omega(Y_1, \dots, [X, Y_i], \dots, Y_k) \quad (15.3.6)$$

**Exercise 15.27.** Prove that  $L_X(L_Y\omega) - L_Y(L_X\omega) = L_{[X,Y]}\omega$ .

**Exercise 15.28.** For the Lie derivative, are there analogues of the equalities in Exercises 14.51, 14.52 and 14.53?

**Exercise 15.29.** Suppose  $F: M \rightarrow N$  is a map, and  $X$  and  $Y$  are  $F$ -related tangent fields (see Exercise 15.3). Prove that  $L_X(F^*\omega) = F^*(L_Y\omega)$ .

## Derivation

Both the exterior derivative and the Lie derivative are derivations. Similar to the bracket of tangent fields, we can use bracket to produce new derivations from known derivations.

**Proposition 15.3.1.** *If  $D_1, D_2$  are derivations of graded algebra  $A$  of degrees  $p$  and  $q$ , then the bracket  $[D_1, D_2] = D_1 D_2 - (-1)^{pq} D_2 D_1$  is a derivation of degree  $p + q$ .*

*Proof.* The bracket  $[D_1, D_2]$  clearly takes  $A_k$  to  $A_{k+p+q}$ . Moreover, for  $a \in A_k$  and

$b \in A_l$ , we have

$$\begin{aligned}
 [D_1, D_2](ab) &= D_1(D_2(ab)) - (-1)^{pq} D_2(D_1(ab)) \\
 &= D_1((D_2a)b + (-1)^{qk} a(D_2b)) - (-1)^{pq} D_2((D_1a)b + (-1)^{pk} a(D_1b)) \\
 &= (D_1D_2a)b + (-1)^{p(q+k)} (D_2a)(D_1b) \\
 &\quad + (-1)^{qk} (D_1a)(D_2b) + (-1)^{qk} (-1)^{pk} a(D_1D_2b) \\
 &\quad - (-1)^{pq} (D_2D_1a)b - (-1)^{pq} (-1)^{q(p+k)} (D_1a)(D_2b) \\
 &\quad - (-1)^{pq} (-1)^{pk} (D_2a)(D_1b) - (-1)^{pq} (-1)^{pk} (-1)^{qk} a(D_2D_1b) \\
 &= (D_1D_2a)b + (-1)^{(p+q)k} a(D_1D_2b) \\
 &\quad - (-1)^{pq} (D_2D_1a)b - (-1)^{pq+(p+q)k} a(D_2D_1b) \\
 &= ([D_1, D_2]a)b - (-1)^{(p+q)k} a([D_1, D_2]b). \quad \square
 \end{aligned}$$

In addition to  $d$  and  $L_X$ , we further introduce the *interior product* along a vector field  $X$

$$i_X: \Omega^k M \rightarrow \Omega^{k-1} M, \quad i_X \omega(Y_1, \dots, Y_{k-1}) = \omega(X, Y_1, \dots, Y_{k-1}).$$

By the third part of Exercise 7.64,  $i_X$  is a derivation of degree  $-1$ . The derivation is determined by the following values at the generators of  $\Omega M$

$$i_X f = 0, \quad i_X df = X(f).$$

The first is due to  $i_X f \in \Omega^{-1} M = 0$ . The second follows from the definition of  $i_X$ , which says that  $i_X \omega = \omega(X)$  for any 1-form  $\omega$ .

Exercise 15.27 shows that  $[L_X, L_Y] = L_{[X, Y]}$ . Exercise 15.31 gives  $[i_X, i_Y]$ .

We have  $[d, d] = dd - (-1)^{1 \cdot 1} dd = 2d^2$ , so that  $d^2$  is a derivative. Then the second property in the definition of  $d$  says that  $d^2(f) = 0$  and  $d^2(df) = 0$ . By Proposition 15.2.3, therefore, we get  $d^2 = 0$  on the whole  $\Omega M$ . This gives an alternative proof of Proposition 15.2.5.

The bracket  $[d, i_X] = di_X - (-1)^{1 \cdot (-1)} i_X d = di_X + i_X d$  has the following values

$$\begin{aligned}
 [d, i_X]f &= di_X f + i_X df = d0 + X(f) = L_X f, \\
 [d, i_X]df &= di_X df + i_X d(df) = d(X(f)) + i_X 0 = L_X df.
 \end{aligned}$$

By Proposition 15.2.3, we get *Cartan's formula*

$$L_X = [d, i_X] = di_X + i_X d. \quad (15.3.7)$$

**Exercise 15.30.** Let  $F: M \rightarrow N$  be a map. For any  $X \in T_x M$  and  $\omega \in \Lambda T_{F(x)}^* N$ , prove that  $i_X F^* \omega = F^* i_{F_* X} \omega$ . The formula can be compared with Exercise ??, but we do not need fields to define  $i_X \omega$ .

**Exercise 15.31.** Prove  $[i_X, i_Y] = 0$ . In fact, any derivation of degree  $-2$  on  $\Omega M$  is 0.

**Exercise 15.32.** Prove  $[d, L_X] = 0$  and  $[i_X, L_Y] = i_{[X, Y]}$ .

**Exercise 15.33.** Prove that the bracket satisfies the Jacobi identity

$$[D_1, [D_2, D_3]] + [D_3, [D_1, D_2]] + [D_2, [D_3, D_1]] = 0.$$

Cartan's formula (15.3.7) can be combined with (15.3.6) to inductively derive the formula of  $d\omega$  as a multilinear alternating function. For example, for  $\omega \in \Omega^1 M$ , we have

$$\begin{aligned} d\omega(X, Y) &= i_X d\omega(Y) = L_X \omega(Y) - di_X \omega(Y) \\ &= [L_X(\omega(Y)) - \omega(L_X Y)] - d(\omega(X))(Y) \\ &= X(\omega(Y)) - \omega([X, Y]) - Y(\omega(X)) \\ &= X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y]). \end{aligned}$$

Here we used the Leibniz rule to calculate  $L_X \omega(Y)$ .

**Proposition 15.3.2.** For  $\omega \in \Omega^k M$ , we have

$$\begin{aligned} d\omega(X_0, X_1, \dots, X_k) &= \sum_{i=0}^k (-1)^i X_i(\omega(X_0, \dots, \hat{X}_i, \dots, X_k)) \\ &\quad + \sum_{0 \leq i < j \leq k} (-1)^{i+j} \omega([X_i, X_j], X_0, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_k). \end{aligned}$$

*Proof.* The formula is already verified for  $k = 1$ . The following is the inductive

verification

$$\begin{aligned}
d\omega(X_0, X_1, \dots, X_k) &= i_{X_0} d\omega(X_1, \dots, X_k) \\
&= L_{X_0} \omega(X_1, \dots, X_k) - di_{X_0} \omega(X_1, \dots, X_k) \\
&= L_{X_0} (\omega(X_1, \dots, X_k)) - \sum_{i=1}^k \omega(X_1, \dots, L_{X_0} X_i, \dots, X_k) \\
&\quad - \sum_{i=1}^k (-1)^{i-1} X_i (i_{X_0} \omega(X_1, \dots, \hat{X}_i, \dots, X_k)) \\
&\quad - \sum_{1 \leq i < j \leq k} (-1)^{i+j} i_{X_0} \omega([X_i, X_j], X_1, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_k) \\
&= X_0 (\omega(X_1, \dots, X_k)) - \sum_{i=1}^k \omega(X_1, \dots, [X_0, X_i], \dots, X_k) \\
&\quad - \sum_{i=1}^k (-1)^{i-1} X_i (\omega(X_0, X_1, \dots, \hat{X}_i, \dots, X_k)) \\
&\quad - \sum_{1 \leq i < j \leq k} (-1)^{i+j} \omega(X_0, [X_i, X_j], X_1, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_k) \\
&= \sum_{i=0}^k (-1)^i X_i (\omega(X_0, \dots, \hat{X}_i, \dots, X_k)) \\
&\quad + \sum_{0 \leq i < j \leq k} (-1)^{i+j} \omega([X_i, X_j], X_0, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_k). \quad \square
\end{aligned}$$

**Exercise 15.34.** Use Exercises 14.52 and 15.4 to extend the last part of Exercise 14.52: For a family  $\omega_t \in \Omega^k M$  of differential forms, we have

$$d\left(\frac{d}{dt}\bigg|_{t=0} \omega_t\right) = \frac{d}{dt}\bigg|_{t=0} d\omega_t, \quad d\left(\int_a^b \omega_t dt\right) = \int_a^b (d\omega_t) dt.$$

## 15.4 Integration

The integration of a function along a submanifold of  $\mathbb{R}^N$  makes use of the natural volume measure on the manifold. The integration of a function on a differentiable manifold should be with respect to a measure on the manifold that is differentiable in some sense. This means that the measure should be linearly approximated at  $x$  by a “linear measure” on  $T_x M$ . Here the linearity of the measure means translation invariance. The whole measure on  $M$  is then given by choosing one translation invariant measure  $\mu_x$  on  $T_x M$  for each  $x \in M$ . In other words, a measure on  $M$  can be regarded as a “measure field”.

**Example 15.4.1.** An increasing function  $\alpha(t)$  induces a measure  $\mu_\alpha$  on the manifold  $\mathbb{R}$ . Assume  $\alpha$  is continuously differentiable. We have  $\mu_\alpha(t, t+\Delta t) = \alpha(t+\Delta t) - \alpha(t) = \alpha'(t^*) \Delta t$  for some  $t^* \in (t, t+\Delta t)$ . Since  $\alpha'(t^*)$  converges to  $\alpha'(t_0)$  as  $t \rightarrow t_0$  and  $\Delta t \rightarrow 0$ , the measure

is approximated by the translation invariant measure  $\mu'(t, t+v) = \alpha'(t_0)(v)$  near  $t_0$ . The measure  $\mu'$  is the  $\alpha'(t_0)$  multiple of the standard Lebesgue measure on  $\mathbb{R}$ .

**Exercise 15.35.** Suppose the linear approximation of a measure on  $\mathbb{R}$  at a point  $x$  is multiplying the usual Lebesgue measure by  $a(x)$ . What should be the measure?

**Exercise 15.36.** Show that a continuously differentiable increasing function  $\alpha(t)$  such that  $\alpha(t+2\pi) - \alpha(t)$  is independent of  $t$  gives a measure on the circle  $S^1$ . What is the linear measure approximating this measure?

## Differentiable Measure on Manifold

Theorem 7.4.2 gives a one-to-one correspondence between translation invariant measures on  $V$  and elements in  $|\Lambda^n V^*| - 0$ . This leads to the following linear approximation of measures on manifolds.

**Definition 15.4.1.** A *differentiable measure* on an  $n$ -dimensional manifold  $M$  assigns an element  $|\omega|(x) \in |\Lambda_x^n T^*M| - 0$  to each  $x \in M$ .

A (local) diffeomorphism  $F: M \rightarrow N$  induces an isomorphism of cotangent spaces and the associated exterior algebras. This further induces the pullback  $F^*: |\Lambda_{F(x)}^n T^*N| - 0 \rightarrow |\Lambda_x^n T^*M| - 0$ . In particular, with respect to any chart  $\sigma: U \rightarrow M$  pulls  $|\omega|$ , we have

$$|\omega| = g|du_1 \wedge \cdots \wedge du_n|, \quad g > 0. \quad (15.4.1)$$

The differentiability of the measure  $|\omega|$  means the differentiability of the function  $g$ .

Let  $\{\sigma_i: U_i \rightarrow M\}$  be a (finite) atlas of  $M$ . Choose (Lebesgue or Borel) measurable subsets  $B_i \subset U_i$ , such that  $M = \cup \sigma_i(B_i)$  and the overlappings between  $\sigma_i(B_i)$  have lower dimension. Then  $|\omega| = g_i|du_1 \wedge \cdots \wedge du_n|$  on  $\sigma_i(B_i)$ , and we define the integral of a function  $f$  on  $M$  with respect to the measure  $|\omega|$  by

$$\int_M f|\omega| = \sum_i \int_{B_i} f(\sigma_i(\vec{u}))g_i(\vec{u})du_1 \cdots du_n.$$

This generalizes the “integral of first type” in multivariable calculus. Similar to the argument (13.2.3), we can show that the integral is independent of the choice of atlas.

A manifold  $M$  is *Riemannian* if the tangent space  $T_x M$  has inner product for each  $x \in M$ . The inner product induces the unique translation invariant measure  $|\omega|$  on  $T_x M$ , such that any unit cube has volume 1. By Theorem 7.4.2 and Proposition 7.4.3, the function  $g$  in the formula (15.4.1) may be calculated from the volume of the parallelootope spanned by the standard basis  $\{\partial_{u_1}, \dots, \partial_{u_n}\}$  of  $T_x M$

$$\begin{aligned} \|\partial_{u_1} \wedge \cdots \wedge \partial_{u_n}\|_2 &= |\omega(\partial_{u_1} \wedge \cdots \wedge \partial_{u_n})| \\ &= g|\langle \partial_{u_1} \wedge \cdots \wedge \partial_{u_n}, du_1 \wedge \cdots \wedge du_n \rangle| = g. \end{aligned}$$



The measure  $|du_1 \wedge \cdots \wedge du_n|$  in (15.4.1) is the standard Lebesgue measure on  $\mathbb{R}^n$ , and is usually denoted by  $du_1 \cdots du_n$  in multivariable calculus. Therefore

$$|\omega| = \|\partial_{u_1} \wedge \cdots \wedge \partial_{u_n}\|_2 du_1 \cdots du_n. \quad (15.4.2)$$

For the special case that  $M$  is a submanifold of  $\mathbb{R}^N$ , the manifold is Riemannian by inheriting the dot product of the Euclidean space. The inner product on the tangent space comes from the embedding

$$\begin{aligned} T_x M \subset T_x \mathbb{R}^N &\cong \mathbb{R}^N : \partial_{u_i} \mapsto \frac{\partial x_1}{\partial u_i} \partial_{x_1} + \cdots + \frac{\partial x_N}{\partial u_i} \partial_{x_N} \\ &\leftrightarrow \sigma_{u_i} = \left( \frac{\partial x_1}{\partial u_i}, \dots, \frac{\partial x_N}{\partial u_i} \right). \end{aligned}$$

Then (15.4.2) becomes the formula (13.3.1) for the volume unit.

### Volume Form

Suppose  $M$  is oriented. Then at each  $x \in M$ , the differentiable measure  $|\omega(x)|$  has a preferred choice  $\omega(x) \in o_x^*$  in the dual orientation component  $o_x^* \subset \Lambda^n T_x^* M - 0$ . This is equivalent to  $\omega(x)(X_1, \dots, X_n) > 0$  for any compatibly oriented basis  $\{X_1, \dots, X_n\}$  of  $T_x M$ . Under an orientation compatible chart  $\sigma: U \rightarrow M$ , this also means taking off the “absolute value” from (15.4.1)

$$\omega = g du_1 \wedge \cdots \wedge du_n, \quad g > 0. \quad (15.4.3)$$

**Proposition 15.4.2.** *An  $n$ -dimensional manifold  $M$  is orientable if and only if there is a nowhere vanishing  $n$ -form  $\omega \in \Omega^n M$ .*

The nowhere vanishing top dimensional form is called a *volume form* because for the compatible orientation given by the proposition, we always have (15.4.3).

*Proof.* Suppose  $\omega$  is a volume form on  $M$ . For any  $x \in M$ , let  $\sigma: U \rightarrow M$  be a chart containing  $x$ . By restricting  $\sigma$  to an open subset of  $U$ , we may assume that  $U$  is connected. Then  $\omega = g(\vec{u}) du_1 \wedge \cdots \wedge du_n$  on  $\sigma(U)$  for a non-vanishing continuous function  $g$  on  $U$ . Since  $U$  is connected, we have either  $g > 0$  everywhere on  $U$ , or  $g < 0$  everywhere on  $U$ . In the first case, we say  $\sigma$  is orientation compatible with  $\omega$ . In the second case, we let  $J(u_1, u_2, \dots, u_n) = (-u_1, u_2, \dots, u_n)$  and find that  $\sigma \circ J$  is orientation compatible with  $\omega$

$$\omega = -g(J(\vec{u})) d(-u_1) \wedge du_2 \wedge \cdots \wedge du_n, \quad -g(J(\vec{u})) > 0.$$

This proves that  $M$  is covered by an atlas in which every chart is orientation compatible with  $\omega$ .

Now for any two charts  $\sigma: U \rightarrow M$  and  $\tau: V \rightarrow M$  that are orientation compatible with  $\omega$ , we have

$$\omega = g du_1 \wedge \cdots \wedge du_n = h dv_1 \wedge \cdots \wedge dv_n, \quad g, h > 0.$$

By Example 15.2.3, the transition  $\varphi = \tau^{-1} \circ \sigma$  satisfies  $\det \varphi' = \frac{g}{h} > 0$ . So the derivatives of transition maps between orientation compatible charts have positive determinants. By Proposition 14.5.2, the orientation compatible charts give an orientation of the manifold.

Conversely, an oriented manifold is covered by an atlas satisfying the property in Proposition 14.5.2. Each chart  $\sigma_i: U_i \rightarrow M_i \subset M$  in the atlas has a standard volume form  $\omega_i = du_1 \wedge \cdots \wedge du_n$  on  $M_i$ . Moreover, we have  $\omega_j = h_{ji}\omega_i$  with  $h_{ji} = \det \varphi'_{ji} > 0$  on the overlapping  $M_i \cap M_j$ . The problem is to patch all the volume form pieces  $\omega_i$  together to form a volume form on the whole  $M$ . This can be achieved by using the *partition of unity*.  $\square$

Given one volume form  $\omega$ , any  $n$ -form is given by  $f\omega$  for a function  $f$ . Therefore we get a one-to-one correspondence between  $C^r(M)$  and  $\Omega^n M$ . In particular, any other volume form on  $M$  is  $f\omega$  for a nowhere vanishing function  $f$ .

If  $\sigma$  is an orientation compatible chart of an oriented Riemannian manifold, then (15.4.2) becomes

$$\omega = \|\partial_{u_1} \wedge \cdots \wedge \partial_{u_n}\|_2 du_1 \wedge \cdots \wedge du_n.$$

This is the unique unit length vector in the dual orientation component  $o_x^* \subset \Lambda^n T_x^* M$ .

**Example 15.4.2.** Let  $M$  be the level manifold  $g(x_0, \dots, x_n) = c$  in  $\mathbb{R}^{n+1}$ , with the orientation induced from the normal vector  $\vec{n} = \frac{\nabla g}{\|\nabla g\|_2}$ . Since  $T_x^* M$  is the orthogonal complement of  $dg = g_{x_0} dx_0 + \cdots + g_{x_n} dx_n$ , we find that  $\Lambda^n T_x^* M \cong \mathbb{R}$  is spanned by

$$(dg)^* = \sum_{i=0}^n g_{x_i} (dx_i)^* = \sum_{i=0}^n (-1)^i g_{x_i} dx_0 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n.$$

For  $X_1, \dots, X_n \in T_{\vec{x}} M$ , we have

$$(dg \wedge (dg)^*)(\vec{n}, X_1, \dots, X_n) = dg(\vec{n}) (dg)^*(X_1, \dots, X_n) = \|\nabla g\|_2 (dg)^*(X_1, \dots, X_n).$$

This implies

$$\begin{aligned} & \{X_1, \dots, X_n\} \text{ is positively oriented in } T_{\vec{x}} M \\ \iff & \{\vec{n}, X_1, \dots, X_n\} \text{ is positively oriented in } \mathbb{R}^{n+1} \\ \iff & (dg)^*(X_1, \dots, X_n) > 0. \end{aligned}$$

Therefore  $(dg)^*$  lies in the dual orientation component  $o_M^* \subset \Lambda^n T_x^* M$  induced by the normal vector. Then we get the volume form as the unique unit length vector in the dual orientation component

$$\omega = \frac{(dg)^*}{\|(dg)^*\|_2} = \frac{\sum_{i=0}^n (-1)^i g_{x_i} dx_0 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n}{\sqrt{g_{x_0}^2 + \cdots + g_{x_n}^2}}. \quad (15.4.4)$$

The formula generalises Exercises 13.46 and 13.52.

The sphere of radius  $R$  in  $\mathbb{R}^{n+1}$  is given by  $g(\vec{x}) = x_0^2 + \cdots + x_n^2 = R^2$ . By  $g_{x_i} = 2x_i$  and  $\|(dg)^*\|_2 = \|\nabla g\|_2 = \|2\vec{x}\|_2 = 2R$ , we get the corresponding volume form

$$\omega = \frac{1}{R} \sum_{i=0}^n (-1)^i x_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n.$$

**Exercise 15.37.** Suppose  $M$  is the submanifold in  $\mathbb{R}^{n+2}$  given by  $g_1(\vec{x}) = g_2(\vec{x}) = 0$ , such that  $dg_1 \wedge dg_2 \neq 0$  along  $M$ . Describe a natural orientation on  $M$  and find a formula for the corresponding volume form.

**Exercise 15.38.** Suppose  $M$  is a submanifold of  $\mathbb{R}^N$  with orientation compatible parameterisation  $\sigma: U \subset \mathbb{R}^k \rightarrow M \subset \mathbb{R}^N$ . Prove that the corresponding volume form is given by

$$\omega = \frac{\sum \det \frac{\partial(x_{i_1}, \dots, x_{i_n})}{\partial(u_1, \dots, u_n)} dx_{i_1} \wedge \cdots \wedge dx_{i_n}}{\sqrt{\sum \left( \det \frac{\partial(x_{i_1}, \dots, x_{i_n})}{\partial(u_1, \dots, u_n)} \right)^2}}.$$

## Partition of Unity

An *open cover* of a manifold  $M$  is a collection  $\{M_i\}$  of open subsets satisfying  $M = \cup M_i$ . A *partition of unity* subordinate to the open cover is a collection of (continuously differentiable) functions  $\alpha_i$  satisfying

1. *Non-negative:*  $\alpha_i \geq 0$  everywhere.
2. *Support:*  $\alpha_i = 0$  on an open subset containing  $M - M_i$ .
3. *Locally finite:* Any  $x \in M$  has a neighborhood  $B$ , such that  $\alpha_i = 0$  on  $B$  for all but finitely many  $i$ .
4. *Unity:*  $\sum_i \alpha_i = 1$ .

Define the *support* of a function  $\alpha$  on  $M$  to be the closure of the places where the function is nonzero

$$\text{supp}(\alpha) = \overline{\{x \in M: \alpha(x) \neq 0\}}.$$

Then the second condition means that  $\text{supp}(\alpha_i) \subset M_i$ , or  $\alpha_i$  is *supported on*  $M_i$ . The concept can be easily extended to tangent fields and differential forms. We also note that the third property says that the sum  $\sum_i \alpha_i(x)$  is always finite near any point. This implies that  $\sum \alpha_i$  is continuous or differentiable if each  $\alpha_i$  is continuous or differentiable.

**Proposition 15.4.3.** *For any open cover of a manifold, there is a partition of unity  $\{\alpha_i\}$ , such that each  $\text{supp}(\alpha_i)$  is contained in a subset in the open cover.*

*Proof.* The manifolds are assumed to be Hausdorff and paracompact. The topological properties imply that any open cover has a locally finite refinement. This implies

that there is an atlas  $\{\sigma_i: U_i \rightarrow M_i = \sigma_i(U_i)\}$ , such that each  $M_i$  is contained in a subset in the open cover, and any  $x \in M$  has an open neighborhood  $B$ , such that  $B \cap M_i = \emptyset$  for all but finitely many  $i$ . The topological properties further imply that the open cover  $\{M_i\}$  has a “shrinking”, which is another open cover  $\{M'_i\}$  such that  $\overline{M'_i} \subset M_i$ . Applying the argument again, we get a further “shrinking”  $\{M''_i\}$  that covers  $M$  and satisfies

$$\overline{M''_i} \subset M'_i \subset \overline{M'_i} \subset M_i.$$

The corresponding open subsets  $U''_i = \sigma_i^{-1}(M''_i)$  and  $U'_i = \sigma_i^{-1}(M'_i)$  satisfy

$$\overline{U''_i} \subset U'_i \subset \overline{U'_i} \subset U_i \subset \mathbb{R}^n.$$

We can find a continuously differentiable (even smooth) function  $\beta_i$  on  $U_i$ , such that  $\beta_i > 0$  on  $U''_i$  and  $\beta_i = 0$  on  $U_i - U'_i$ . This implies that the function

$$\tilde{\beta}_i = \begin{cases} \beta_i \circ \sigma_i^{-1}, & \text{on } M_i \\ 0, & \text{on } M - M_i \end{cases}$$

is continuously differentiable on  $M$ , with  $\tilde{\beta}_i > 0$  on  $M''_i$  and  $\text{supp}(\tilde{\beta}_i) \subset \overline{M'_i} \subset M_i$ .

Since the collection  $\{M_i\}$  is locally finite, the sum  $\sum_i \tilde{\beta}_i$  is finite near any point and is therefore still continuously differentiable. Since  $\beta_i > 0$  on  $M''_i$  and  $\{M''_i\}$  covers  $M$ , we have  $\sum_i \tilde{\beta}_i > 0$ . Then

$$\alpha_i = \frac{\tilde{\beta}_i}{\sum_j \tilde{\beta}_j}$$

is a continuously differentiable partition of unity satisfying  $\text{supp}(\alpha_i) \subset M_i$ .  $\square$

Now we use partition of unity to finish the proof of Proposition 15.4.2. We may further assume that the atlas  $\{\sigma_i: U_i \rightarrow M_i\}$  in the proof is locally finite, and  $\{\alpha_i\}$  is a partition of unity satisfying  $\text{supp}(\alpha_i) \subset M_i$ . Then the extension of  $\alpha_i \omega_i$  on  $M_i$  to  $M$  by assigning  $\alpha_i \omega_i = 0$  on  $M - M_i$  is a differentiable  $n$ -form. The local finiteness of the atlas implies that the sum

$$\omega = \sum_i \alpha_i \omega_i$$

is still a differentiable  $n$ -form on  $M$ . For any  $x \in M$ , there are only finitely many charts  $\sigma_{i_0}, \sigma_{i_1}, \dots, \sigma_{i_l}$ , such that all other  $\alpha_i = 0$  near  $x$ . Then we have

$$\omega = \sum_{j=0}^l \alpha_{i_j} \omega_{i_j} = (1 + h_{i_1 i_0} + \dots + h_{i_l i_0}) \omega_{i_0} \text{ near } x.$$

By  $\omega_{i_0} \neq 0$  and  $h_{i i_0} \geq 0$  near  $x$ , we conclude that  $\omega \neq 0$  near  $x$ . This shows that  $\omega$  is a volume form.

## Integration of Differential Form

Let  $\omega$  be a volume form on an  $n$ -dimensional oriented manifold  $M$ . Then we may define the integral  $\int_M f\omega$  of functions  $f$  with respect to the volume form. Since  $f\omega$  is simply an  $n$ -form, we will actually define the integral  $\int_M \omega$  of any  $n$ -form  $\omega$ .

If the manifold is covered by one chart  $\sigma: U \rightarrow M$  (or  $\omega = 0$  outside  $\sigma(U)$ ), then  $\sigma^*\omega = g du_1 \wedge \cdots \wedge du_n$ , where  $du_1 \wedge \cdots \wedge du_n$  is the standard volume form of  $\mathbb{R}^n$ . We define

$$\int_M \omega = \int_{\sigma(U)} \omega = \int_U \sigma^*\omega = \int_U g du_1 \wedge \cdots \wedge du_n = \int_U g du_1 \cdots du_n.$$

In general, we have an orientation compatible atlas  $\{\sigma_i: U_i \rightarrow M_i \subset M\}$ . By Proposition 15.4.3, there is a partition of unity  $\alpha_i$  satisfying  $\text{supp}(\alpha_i) \subset M_i$ . Then we define

$$\int_M \omega = \sum_i \int_M \alpha_i \omega = \sum_i \int_{M_i} \alpha_i \omega = \sum_i \int_{U_i} \sigma_i^*(\alpha_i \omega).$$

To keep the sum on the right to be finite, we additionally assume that  $M$  is a *compact* manifold, so we may choose an atlas with only finitely many charts.

To verify that the definition is independent of the choice of the atlas and the partition of unity, consider to atlases  $\{\sigma_i: U_i \rightarrow M_i \subset M\}$  and  $\{\tau_j: V_j \rightarrow N_j \subset M\}$ , with corresponding partitions of unity  $\alpha_i$  and  $\beta_j$ . Then we get a refinement of the first atlas by

$$\sigma_{ij} = \sigma_i|_{U_{ij}}: U_{ij} = \sigma_i^{-1}(M_i \cap N_j) \rightarrow M_i \cap N_j \subset M,$$

and a refinement of the second atlas by

$$\tau_{ji} = \tau_j|_{V_{ji}}: V_{ji} = \tau_j^{-1}(M_i \cap N_j) \rightarrow M_i \cap N_j \subset M.$$

Moreover, the functions  $\alpha_i \beta_j$  form a partition of unity for both refinements.

The definition of the integral according to the atlas  $\{\sigma_i\}$  is

$$\begin{aligned} \int_M \omega &= \sum_i \int_{U_i} \sigma_i^*(\alpha_i \omega) = \sum_i \int_{U_i} \sigma_i^* \left( \sum_j \beta_j \alpha_i \omega \right) \\ &= \sum_{ij} \int_{U_i} \sigma_i^*(\beta_j \alpha_i \omega) = \sum_{ij} \int_{U_{ij}} \sigma_{ij}^*(\beta_j \alpha_i \omega), \end{aligned}$$

where the last equality is due to  $\alpha_i \beta_j = 0$  out of  $M_i \cap N_j$ . Similarly, the definition of the integral according to  $\{\tau_j\}$  is

$$\int_M \omega = \sum_{ij} \int_{V_{ji}} \tau_{ji}^*(\alpha_i \beta_j \omega).$$

We note that  $\sigma = \sigma_{ij}: U = U_{ij} \rightarrow M$  and  $\tau = \tau_{ji}: V = V_{ji} \rightarrow M$  are two orientation compatible charts, with  $\sigma(U) = \tau(V) = M_i \cap N_j$ . So the problem is reduced to

$$\int_U \sigma^* \omega = \int_V \tau^* \omega,$$

for any  $n$ -form  $\omega$  on  $M_i \cap N_j$ .

Let  $\varphi = \tau^{-1} \circ \sigma: U \rightarrow V$  be the transition map between the two orientation compatible charts. By Example 15.2.3, if

$$\tau^* \omega = g(\vec{v}) dv_1 \wedge \cdots \wedge dv_n,$$

then we have

$$\sigma^* \omega = g(\varphi(\vec{u})) (\det \varphi'(\vec{u})) du_1 \wedge \cdots \wedge du_n.$$

Therefore

$$\begin{aligned} \int_V \tau^* \omega &= \int_V g(\vec{v}) dv_1 \cdots dv_n \\ &= \int_U g(\varphi(\vec{u})) |\det \varphi'(\vec{u})| du_1 \cdots du_n \\ &= \int_U g(\varphi(\vec{u})) (\det \varphi'(\vec{u})) du_1 \cdots du_n = \int_U \sigma^* \omega. \end{aligned}$$

Here the second equality is the change of variable formula for the integral on Euclidean space (Theorem 12.4.5), and the third equality is due to the fact that the orientation compatibility of  $\sigma$  and  $\tau$  implies  $\det \varphi'(u) > 0$ .

**Example 15.4.3.** Suppose the surface in Example 15.2.2 is oriented. Then for an orientable compatible atlas  $\sigma_i(u, v)$  and the corresponding partition of unity  $\alpha_i(u, v)$ , the integral of the 2-form  $\omega = f dy \wedge dz + g dz \wedge dx + h dx \wedge dy$  along the surface is

$$\int_M \omega = \sum_i \int_{U_i} \alpha_i(u, v) \left[ f(\sigma_i(u, v)) \frac{\partial(y, z)}{\partial(u, v)} + g(\sigma_i(u, v)) \frac{\partial(z, x)}{\partial(u, v)} + h(\sigma_i(u, v)) \frac{\partial(x, y)}{\partial(u, v)} \right] dudv.$$

In practice, we do not use the partition of unity to break the integral into pieces. Instead, we use subsets  $B_i \subset U_i$ , such that  $\sigma_i(B_i)$  covers  $M$  and the dimension of the overlappings  $\leq 1$ . Then the integral is

$$\int_M \omega = \sum_i \int_{B_i} \left[ f(\sigma_i(u, v)) \frac{\partial(y, z)}{\partial(u, v)} + g(\sigma_i(u, v)) \frac{\partial(z, x)}{\partial(u, v)} + h(\sigma_i(u, v)) \frac{\partial(x, y)}{\partial(u, v)} \right] dudv.$$

The expression can be viewed as taking  $\alpha_i = \chi_{\sigma_i(B_i)}$  as the non-differentiable partition of unity.

## Stokes' Theorem

**Theorem 15.4.4 (Stokes' Theorem).** Suppose  $M$  is a  $n$ -dimensional compact oriented manifold with boundary. Then for any  $\omega \in \Omega^{n-1}M$ , we have

$$\int_M d\omega = \int_{\partial M} \omega.$$

*Proof.* Let  $\alpha_i$  be a partition of unity corresponding to an atlas. The following shows that it is sufficient to prove  $\int_M d(\alpha_i \omega) = \int_{\partial M} \alpha_i \omega$ .

$$\begin{aligned} \int_{\partial M} \omega &= \sum \int_{\partial M} \alpha_i \omega = \sum \int_M d(\alpha_i \omega) = \sum \int_M (d\alpha_i \wedge \omega + \alpha_i d\omega) \\ &= \int_M d\left(\sum \alpha_i\right) \wedge \omega + \int_M \left(\sum \alpha_i\right) d\omega = \int_M d1 \wedge \omega + \int_M 1 d\omega = \int_M d\omega. \end{aligned}$$

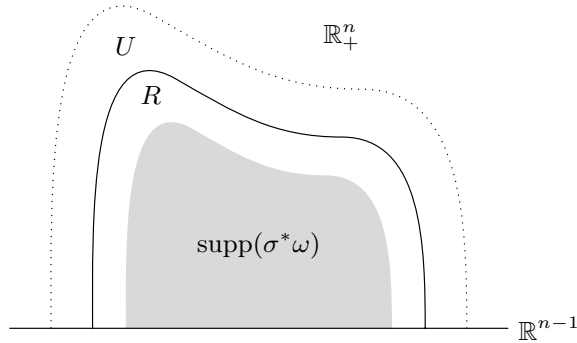
Since  $\alpha_i \omega$  is contained in a chart, we only need to prove the equality for the case  $\omega$  vanishes outside a chart  $\sigma: U \subset \mathbb{R}_+^n \rightarrow M$ . By the definition of integration and  $\sigma^* d\omega = d\sigma^* \omega$  (see Proposition 15.2.4), the problem is reduced to

$$\int_U d\sigma^* \omega = \int_{U \cap \mathbb{R}^{n-1}} \sigma^* \omega.$$

Here  $U$  has the standard orientation of  $\mathbb{R}^n$ , and  $\mathbb{R}^{n-1}$  has the induced orientation given by Definition 14.5.4. By the discussion following the definition,  $\mathbb{R}^{n-1}$  is negatively oriented if  $n$  is odd, and is positively oriented if  $n$  is even.

Since the support of  $\sigma^* \omega$  is a closed subset of a compact manifold, by Exercises 6.65 and 14.29, the support is a compact subset of the open subset  $U$  of the half space  $\mathbb{R}_+^n$ . This implies that there is a nice region  $R$  satisfying  $\text{supp}(\sigma^* \omega) \subset R \subset U$ . The boundary  $\partial R$  of the region has two parts. The first part  $R \cap \mathbb{R}^{n-1}$  lies in  $\mathbb{R}^{n-1}$ , and we have  $\sigma^* \omega = 0$  on the second part  $\partial R - \mathbb{R}^{n-1}$ . Therefore the equality we wish to prove is

$$\int_R d\sigma^* \omega = \int_{\partial R} \sigma^* \omega = \int_{R \cap \mathbb{R}^{n-1}} \sigma^* \omega.$$



**Figure 15.4.1.** Nice region  $R$  between  $U$  and  $\text{supp}(\sigma^* \omega)$ .

Let

$$\sigma^* \omega = \sum_{i=1}^n f_i(\vec{u}) du_1 \wedge \cdots \wedge \widehat{du_i} \wedge \cdots \wedge du_n.$$

Then

$$\begin{aligned} d\sigma^*\omega &= \sum_{i=1}^n \frac{\partial f_i}{\partial u_i} du_i \wedge du_1 \wedge \cdots \wedge \widehat{du_i} \wedge \cdots \wedge du_n \\ &= \sum_{i=1}^n (-1)^{i-1} \frac{\partial f_i}{\partial u_i} du_1 \wedge \cdots \wedge du_n. \end{aligned}$$

So the equality we wish to prove is

$$\int_R \sum_{i=1}^n (-1)^{i-1} \frac{\partial f_i}{\partial u_i} du_1 \wedge \cdots \wedge du_n = \int_{\partial R} \sum_{i=1}^n f_i du_1 \wedge \cdots \wedge \widehat{du_i} \wedge \cdots \wedge du_n.$$

This is exactly Gauss' Theorem 13.6.2.  $\square$

**Example 15.4.4.** A subset  $Y \subset X$  is a *retract* if there is a map  $F: X \rightarrow Y$ , such that  $F(y) = y$ . The map  $F$  is a *retraction*. We claim that for a compact oriented manifold  $M$ , its boundary  $\partial M$  cannot be a (continuously differentiable) retract of  $M$ . In particular, the sphere is not a retract of the ball.

Suppose there is a retract  $F: M \rightarrow \partial M$ . Since  $\partial M$  is also oriented, by Proposition 15.4.2,  $\partial M$  has a volume form  $\omega \in \Omega^{n-1}\partial M$ . Then

$$\int_{\partial M} \omega = \int_{\partial M} F^*\omega = \int_M dF^*\omega = \int_M F^*d\omega = \int_M F^*0 = 0.$$

Here the first equality is due to  $F|_{\partial M}$  being the identity map, the second equality is by Stokes' Theorem, and the last equality is due to  $d\omega \in \Omega^n\partial M = 0$  by  $n > \dim \partial M$ . On the other hand, the integral  $\int_{\partial M} \omega$  of the volume form should always be positive. So we get a contradiction.

**Exercise 15.39 (Brouwer's Fixed Point Theorem).** Prove that if  $F: B^n \rightarrow B^n$  is a map without fixed point, then there is a retract of  $B^n$  to the boundary  $S^{n-1}$ . Then show that any map of  $B^n$  to itself has a fixed point.

## 15.5 Homotopy

### Closed Form and Exact Form

A differential form  $\omega$  is *closed* if  $d\omega = 0$ . It is *exact* if  $\omega = d\varphi$  for a differential form  $\varphi$ , called a *potential* of  $\omega$ . By  $d^2 = 0$ , exact forms must be closed.

A function  $f \in \Omega^1 M$  is closed if  $df = 0$ . This means that  $f$  is a constant. Since  $\Omega^{-1}M = 0$ , the only exact function is the zero function.

A 1-form  $\omega = f_1 dx_1 + \cdots + f_n dx_n$  on an open subset  $U \subset \mathbb{R}^n$  is closed if and only if

$$\frac{\partial f_j}{\partial x_i} = \frac{\partial f_i}{\partial x_j}.$$

Theorem 13.5.3 says that  $\omega$  is exact on  $U$  if and only if the integral  $\int_{\gamma} \omega$  along any curve  $\gamma$  in  $U$  depends only on the beginning and end points of  $\gamma$ . This is also



equivalent to  $\int_{\gamma} \omega = 0$  for any closed curve  $\gamma$  in  $U$ . The theorem also says that, if  $U$  satisfies some topological condition, then  $\omega$  is exact if and only if it is closed. Since the topological condition is always satisfied by balls, this means that locally, closedness and exactness are equivalent. Therefore the distinction between the two concepts is global and topological.

The goal of this section is to extend the observations about the closed and exact 1-forms on Euclidean space to  $k$ -forms on manifolds.

**Example 15.5.1.** A 1-form  $\omega = \sum_{i=0}^n f_i dx_i$  is closed on the sphere  $S^n \subset \mathbb{R}^{n+1}$  if and only if the evaluation of  $d\omega$  on tangent vectors of  $S^n$  is always zero. Since  $T_{\vec{x}} S^n = (\mathbb{R}\vec{x})^{\perp}$  is spanned by  $X_i = x_i \partial_{x_0} - x_0 \partial_{x_i}$ ,  $i = 1, \dots, n$ , this means the vanishing of the following

$$\begin{aligned} d\omega(X_i, X_j) &= \sum_{0 \leq k < l \leq n} \left( \frac{\partial f_l}{\partial x_k} - \frac{\partial f_k}{\partial x_l} \right) dx_k \wedge dx_l(X_i, X_j) \\ &= \left( \frac{\partial f_i}{\partial x_0} - \frac{\partial f_0}{\partial x_i} \right) \det \begin{pmatrix} x_i & x_j \\ -x_0 & 0 \end{pmatrix} \\ &\quad + \left( \frac{\partial f_j}{\partial x_0} - \frac{\partial f_0}{\partial x_j} \right) \det \begin{pmatrix} x_i & x_j \\ 0 & -x_0 \end{pmatrix} \\ &\quad + \left( \frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} \right) \det \begin{pmatrix} -x_0 & 0 \\ 0 & -x_0 \end{pmatrix} \\ &= x_0 \left( x_j \frac{\partial f_j}{\partial x_i} - x_j \frac{\partial f_i}{\partial x_j} - x_i \frac{\partial f_j}{\partial x_0} + x_i \frac{\partial f_0}{\partial x_j} + x_0 \frac{\partial f_j}{\partial x_i} - x_0 \frac{\partial f_i}{\partial x_j} \right) \\ &= x_0 \det \begin{pmatrix} x_0 & \partial_0 & f_0 \\ x_i & \partial_i & f_i \\ x_j & \partial_j & f_j \end{pmatrix}. \end{aligned}$$

Of course, there is nothing special about the index 0, and the condition can be rephrased as

$$\det \begin{pmatrix} x_i & \partial_i & f_i \\ x_j & \partial_j & f_j \\ x_k & \partial_k & f_k \end{pmatrix} = 0.$$

**Exercise 15.40.** Prove that if  $\omega$  is closed and  $\rho$  is closed, then  $\omega \wedge \rho$  is closed. Prove that if  $\omega$  is closed and  $\rho$  is exact, then  $\omega \wedge \rho$  is exact.

**Exercise 15.41.** On connected subsets, the potential of an exact 1-form is unique up to adding constants. Do you have the similar statement for the potential of an exact 2-form?

**Exercise 15.42.** The volume form for the sphere in Example 15.4.2 suggests the following differential form

$$\omega = g(\|\vec{x}\|_2) \sum_{i=0}^n (-1)^i x_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n$$

that is invariant with respect to rotations around the origin. Find the condition on the function  $g$  such that  $\omega$  is closed.

## Potential of 1-Form

The following extends Theorem 13.5.3 to manifolds.

**Theorem 15.5.1.** *For a 1-form  $\omega$  on  $M$ , the following are equivalent.*

1. *The integral  $\int_{\gamma} \omega = \int_I \gamma^* \omega$  of  $\omega$  along a curve  $\gamma: I = [a, b] \rightarrow M$  depends only on the beginning  $\gamma(a)$  and end  $\gamma(b)$  of  $\gamma$ .*
2.  *$\omega$  is exact: There is a function  $\varphi$  on  $M$ , such that  $\omega = d\varphi$ .*

Moreover, if any closed curve in  $M$  is the boundary of a map  $\sigma: S \rightarrow M$ , where  $S$  is an orientable surface with circle boundary, then the above is also equivalent to  $d\omega = 0$  (i.e.,  $\omega$  is closed).

*Proof.* Let a curve  $\gamma: I = [a, b] \rightarrow M$  connect  $\gamma(a) = x_0$  to  $\gamma(b) = x$ . Then  $\omega = d\varphi$  implies

$$\varphi(x) - \varphi(x_0) = \int_{\partial I} \varphi \circ \gamma = \int_I d(\varphi \circ \gamma) = \int_I \gamma^* d\varphi = \int_I \gamma^* \omega = \int_{\gamma} \omega. \quad (15.5.1)$$

Here the first two equalities is the reformulation of the Fundamental Theorem of Calculus as Stokes' Theorem (Theorem 15.4.4). The equality clearly shows that  $\int_I \gamma^* \omega$  depends only on the values of  $\varphi$  at the end points  $x_0$  and  $x$  of  $\gamma$ . This proves that the second statement implies the first.

Conversely, suppose the first statement holds. On a connected manifold  $M$ , we fix a point  $x_0 \in M$ , fix a value  $\varphi(x_0)$  (say  $\varphi(x_0) = 0$ ), and then use (15.5.1) to define a function  $\varphi$  on  $M$ . The definition makes use of a curve connecting  $x_0$  to  $x$ , and the first statement means that the definition is independent of the choice of the curve. Next we verify that  $d\varphi = \omega$  at  $x \in M$ . This means that we need to verify  $\langle [\gamma], d_x \varphi \rangle = \langle [\gamma], \omega(x) \rangle$  for any curve  $\gamma: [a, b] \rightarrow M$ ,  $\gamma(0) = x$ . By fixing  $\gamma$  on  $(-\delta, \delta)$  and modifying  $\gamma$  on  $[a, b]$ , we may assume that  $\gamma(a) = x_0$ . Since  $\gamma$  is not changed on  $(-\delta, \delta)$ , the tangent vector  $[\gamma] \in T_x M$  is not changed by the modification. Let  $\gamma^* \omega = f(t)dt$ . Then  $\gamma^* \omega(x) = f(0)dt$ , and we have

$$\begin{aligned} \langle [\gamma], d_x \varphi \rangle &= \left. \frac{d}{dt} \right|_{t=0} \varphi(\gamma(t)) = \left. \frac{d}{dt} \right|_{t=0} \left( \varphi(x_0) + \int_a^t \gamma^* \omega \right) \\ &= \left. \frac{d}{dt} \right|_{t=0} \int_a^t f(u)du = f(0), \\ \langle [\gamma], \omega(x) \rangle &= \langle \gamma_* \partial_t, \omega(x) \rangle = \langle \partial_t, \gamma^* \omega(x) \rangle = \langle \partial_t, f(0)dt \rangle = f(0). \end{aligned}$$

This verifies  $d\varphi = \omega$  at  $x \in M$ .

Finally, we need to show that, under additional topological assumption on  $M$ , the exactness is implied by the weaker closedness. By the assumption, any closed curve  $\gamma$  is the boundary of a map  $\sigma$ . Then by Stokes' Theorem (Theorem 15.4.4),

the integral of a closed 1-form  $\omega$  along an closed curve is

$$\int_{\partial S} \gamma^* \omega = \int_{\partial S} \sigma^* \omega = \int_S d\sigma^* \omega = \int_S \sigma^* d\omega = \int_S \sigma^* 0 = 0.$$

This is equivalent to the first statement.  $\square$

**Example 15.5.2.** For  $n > 1$ , the sphere satisfies the extra condition in Theorem 15.5.1. By Example 15.5.1, a 1-form  $\omega = \sum_{i=0}^n f_i dx_i$  has potential on any sphere entered at the origin if and only if

$$\det \begin{pmatrix} x_i & \partial_i & f_i \\ x_j & \partial_j & f_j \\ x_k & \partial_k & f_k \end{pmatrix} = 0, \text{ for all } i, j, k.$$

**Exercise 15.43.** What is the condition for  $\omega = f dx + g dy$  to have potential on the circle  $S^1 \subset \mathbb{R}^2$ ?

**Exercise 15.44.** Find all the 1-forms  $\omega = f dx + g dy + h dz$  satisfying  $d\omega = 0$  when restricted to any surface submanifold  $x + y + z = c$ . What does this mean for the integral  $\int_{\gamma} \omega$  along curves  $\gamma$ ?

## Homotopy

The key to extending Theorem 15.5.1 to general differential forms is the extension of formula (15.5.1) to the case  $\omega \in \Omega^{k+1}M$  and  $\varphi \in \Omega^k M$ . The problem is that  $\varphi(x) - \varphi(x_0)$  does not make sense for  $k$ -form  $\varphi$ , because the two vectors belong to different vector spaces. So we evaluate (actually integrate) both sides at oriented  $k$ -dimensional manifolds in  $M$ . Specifically, if  $\varphi \in \Omega^k M$ ,  $N$  is a  $k$ -dimensional oriented manifold, and  $F: N \rightarrow M$  is a map, then we have  $\int_N F^* \varphi$ . This extends the special case of  $k = 0$ , when  $\varphi$  is a function,  $N$  is a single point,  $F: N \rightarrow M$  is a point  $x = F(N)$  in  $M$ , and  $\int_N F^* \varphi = \varphi(x)$  is the value of the function at a point. So we replace points  $x_0$  and  $x$  in the formula (15.5.1) by maps (or “super-points”)  $F_0, F: N \rightarrow M$  from oriented manifolds, and replace the curve  $\gamma$  as “moving a super-point”  $F_0$  to “another super-point”  $F$ . The movement of maps is the following definition.

**Definition 15.5.2.** A *homotopy* between two maps  $F_0, F_1: N \rightarrow M$  is a map  $H: [0, 1] \times N \rightarrow M$ , such that  $F_0(x) = H(0, x)$  and  $F_1(x) = H(1, x)$ .

A homotopy is usually only assumed to be continuous. We will always assume that it is sufficiently differentiable. In fact, any continuous homotopy between differentiable maps can be approximated by a differentiable homotopy.

The map  $H$  plays the role of  $\gamma$ . For  $\omega = d\varphi$ , by

$$\partial([0, 1] \times N) = (1 \times N) \cup (-0 \times N) \cup (-[0, 1] \times \partial N),$$

we get the following  $k$ -dimensional analogue of (15.5.1)

$$\begin{aligned}
 \int_N F_1^* \varphi - \int_N F_0^* \varphi &= \int_{(1 \times N) \cup (-0 \times N)} H^* \varphi \\
 &= \int_{\partial([0,1] \times N)} H^* \varphi + \int_{[0,1] \times \partial N} H^* \varphi \\
 &= \int_{[0,1] \times N} dH^* \varphi + \int_{[0,1] \times \partial N} H^* \varphi \\
 &= \int_N \int_0^1 H^*(d\varphi) + \int_{\partial N} \int_0^1 H^* \varphi \\
 &= \int_N \left( \int_0^1 H^*(d\varphi) + d \int_0^1 H^* \varphi \right).
 \end{aligned}$$

Here the third and fifth equalities use Stokes' Theorem (Theorem 15.4.4), and the fourth equality assumes certain version of Fubini Theorem (Theorem 11.3.4). The computation suggests us to introduce an operation of “partial integration along the  $[0, 1]$ -direction”

$$I = \int_0^1 : \Omega^{k+1}([0, 1] \times N) \rightarrow \Omega^k N,$$

such that

$$\int_{[0,1] \times N} \omega = \int_N I(\omega), \quad F_1^* \varphi - F_0^* \varphi = (Id + dI)(H^* \varphi).$$

Now we rigorously carry out the idea. Let  $N$  be a manifold, not necessarily oriented or  $k$ -dimensional. Let  $\omega \in \Omega^{k+1}([0, 1] \times N)$ . We have  $T_{(t,x)}([0, 1] \times N) = \mathbb{R}\partial_t \oplus T_x N$ , which means that a tangent vector of  $[0, 1] \times N$  is of the form  $a\partial_t + X$ , with  $a \in \mathbb{R}$  and  $X \in TM$ . Define  $\lambda \in \Omega^k N$  and  $\mu \in \Omega^{k+1} N$  by

$$\lambda(X_1, \dots, X_k) = \omega(\partial_t, X_1, \dots, X_k), \quad \mu(X_0, \dots, X_k) = \omega(X_0, \dots, X_k),$$

where the same  $X_i \in T_x N$  can be viewed as  $0\partial_t + X_i \in T_{(t,x)}([0, 1] \times N)$  for any  $t \in [0, 1]$ . Since the evaluations of  $\omega$  above has hidden parameter  $t$ ,  $\lambda$  and  $\mu$  are actually families of differential forms parameterised by  $t$ . In a chart of  $N$ , we have

$$\begin{aligned}
 \lambda &= \sum f_{i_1 \dots i_k}(t, \vec{u}) du_{i_1} \wedge \dots \wedge du_{i_k}, \\
 \mu &= \sum g_{i_0 \dots i_k}(t, \vec{u}) du_{i_0} \wedge \dots \wedge du_{i_k}.
 \end{aligned}$$

The definitions of  $\lambda$  and  $\mu$  say that the equality

$$\omega = dt \wedge \lambda + \mu \tag{15.5.2}$$

holds when evaluated at  $\partial_t \wedge X_1 \wedge \dots \wedge X_k$  and  $X_0 \wedge X_1 \wedge \dots \wedge X_k$ . Since  $\Omega^{k+1}([0, 1] \times N)$  consists of linear combinations (with function coefficients) of these vectors, the equality holds on  $\Omega^{k+1}([0, 1] \times N)$ .

For each fixed  $x \in N$ ,  $\lambda(x)$  can be viewed as a curve in the exterior product space

$$t \in [0, 1] \mapsto \lambda_t(x) \in \Lambda^k T_x^* M.$$

Then the integral  $\int_0^1 \lambda_t(x) dt$  makes sense, and is again a vector in  $\Lambda^k T_x^* M$ . By considering all  $x \in N$ , the integral  $\int_0^1 \lambda_t dt$  is a  $k$ -form on  $N$ . This is the operation  $I$  in the following result.

**Lemma 15.5.3.** *Let  $I: \Omega^{k+1}([0, 1] \times N) \rightarrow \Omega^k N$  be given by*

$$I\omega(X_1, \dots, X_k) = \int_0^1 \omega(\partial_t, X_1, \dots, X_k) dt.$$

*Then for  $\omega \in \Omega^{k+1}([0, 1] \times N)$  and the embeddings*

$$i_0(x) = (0, x): N \rightarrow [0, 1] \times N, \quad i_1(x) = (1, x): N \rightarrow [0, 1] \times N,$$

*we have*

$$i_1^* \omega - i_0^* \omega = Id\omega + dI\omega.$$

*Proof.* The  $t$ -parameterised  $\lambda$  can be considered as a form on  $[0, 1] \times N$  or a form on  $N$ . The two respective exterior derivatives are related by

$$\begin{aligned} d_{[0,1] \times N} \lambda &= \sum_{0 \leq i_1 < \dots < i_k \leq n} \frac{\partial f_{i_1 \dots i_k}(t, \vec{u})}{\partial t} dt \wedge du_{i_1} \wedge \dots \wedge du_{i_k} \\ &+ \sum_{0 \leq i_1 < \dots < i_k \leq n} \sum_j \frac{\partial f_{i_1 \dots i_k}(t, \vec{u})}{\partial u_j} du_j \wedge du_{i_1} \wedge \dots \wedge du_{i_k} \\ &= dt \wedge \frac{\partial \lambda}{\partial t} + d_N \lambda. \end{aligned}$$

The same remark applies to  $\mu$ . Then by (15.5.2), we have ( $d\omega = d_{[0,1] \times N} \omega$ ,  $d\lambda = d_N \lambda$  and  $d\mu = d_N \mu$ )

$$d\omega = dt \wedge \left( dt \wedge \frac{\partial \lambda}{\partial t} + d\lambda \right) + dt \wedge \frac{\partial \mu}{\partial t} + d\mu = dt \wedge \left( d\lambda + \frac{\partial \mu}{\partial t} \right) + d\mu.$$

This implies

$$Id\omega = \int_0^1 \left( d\lambda + \frac{\partial \mu}{\partial t} \right) dt = d \left( \int_0^1 \lambda dt \right) + \mu|_{t=1} - \mu|_{t=0} = dI\omega + i_0^* \mu - i_1^* \mu.$$

In the second equality, Exercise 15.34 and classical Fundamental Theorem of Calculus are used.  $\square$

**Exercise 15.45.** For a loop  $\gamma: S^1 \rightarrow \mathbb{R}^2$  and a point  $\vec{a} = (a, b)$  not on the loop, define the *winding number* of  $\gamma$  with respect to  $\vec{a}$

$$\int_{\gamma} \frac{-(y-b)dx + (x-a)dy}{(x-a)^2 + (y-b)^2}.$$

Prove the following properties of winding numbers.

1. If  $\gamma_1$  and  $\gamma_2$  are two homotopic loops in  $\mathbb{R}^2 - \vec{a}$ , then the two winding numbers are the same.
2. If  $\vec{a}$  is moved to another point  $\vec{b}$  without crossing  $\gamma$ , then the winding number is not changed.
3. If  $\vec{a}$  is outside  $\gamma$ , in the sense that  $\vec{a}$  can be moved as far as we like without crossing  $\gamma$ , then the winding number is 0.

What happens to the winding number when  $\vec{a}$  crosses  $\gamma$ ?

**Exercise 15.46.** Suppose  $N$  is an oriented  $k$ -dimensional manifold *without boundary*. Suppose and  $H: [0, 1] \times N \rightarrow M$  is a homotopy between  $F_0, F_1: N \rightarrow M$ . Prove that for any closed  $\omega \in \Omega^k M$ , we have  $\int_N F_1^* \omega = \int_N F_0^* \omega$ .

## Poincaré Lemma

With the help of Lemma 15.5.3, we are ready to extend Theorem 15.5.1 to  $k$ -forms.

A manifold is *contractible* if there is a homotopy  $H: [0, 1] \times M \rightarrow M$ , such that  $H_1 = id$  and  $H_0$  maps  $M$  to a single point. In other words, the identity map is homotopic to the constant map.

**Theorem 15.5.4 (Poincaré Lemma).** *A differential form on a contractible manifold is exact if and only if it is closed.*

*Proof.* We only need to show that  $d\omega = 0$  implies  $\omega = d\varphi$  for another differential form  $\varphi$ . Using Lemma 15.5.3 and the homotopy, we get

$$\omega = H_1^* \omega - H_0^* \omega = i_1^* H^* \omega - i_0^* H^* \omega = IH^* d\omega + dIH^* \omega = dIH^* \omega.$$

This shows that  $\omega = d\varphi$  for  $\varphi = IH^* \omega$ . □

Since a manifold is locally a Euclidean space, which is contractible (see Example 15.5.3), the theorem says that closedness and exactness are locally equivalent.

**Example 15.5.3.** The homotopy  $H(t, \vec{x}) = t\vec{x}: \mathbb{R}^n \rightarrow \mathbb{R}^n$  shows that the Euclidean space is contractible. The  $i$ -th coordinate of the homotopy is  $h_i(t, \vec{x}) = tx_i$ . Therefore  $H^* dx_i =$

$dh_i = x_i dt + t dx_i$ , and

$$\begin{aligned} H^*(g dx_{i_1} \wedge \cdots \wedge dx_{i_k}) &= g(H(t, \vec{x})) dh_{i_1} \wedge \cdots \wedge dh_{i_k} \\ &= g(t\vec{x})(x_{i_1} dt + t dx_{i_1}) \wedge \cdots \wedge (x_{i_k} dt + t dx_{i_k}) \\ &= dt \wedge \left( t^{k-1} g(t\vec{x}) \sum_{p=1}^k (-1)^{p-1} x_{i_p} dx_{i_1} \wedge \cdots \wedge \widehat{dx_{i_p}} \wedge \cdots \wedge dx_{i_k} \right) \\ &\quad + t^k g(t\vec{x}) dx_{i_1} \wedge \cdots \wedge dx_{i_k}, \\ IH^*(g dx_{i_1} \wedge \cdots \wedge dx_{i_k}) &= \left( \int_0^1 t^{k-1} g(t\vec{x}) dt \right) \sum_{p=1}^k (-1)^{p-1} x_{i_p} dx_{i_1} \wedge \cdots \wedge \widehat{dx_{i_p}} \wedge \cdots \wedge dx_{i_k}. \end{aligned}$$

The exterior derivative of a general  $k$ -form  $\omega = \sum a_{i_1 \dots i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$  is given by (15.2.3). The form is closed if and only if

$$\sum_{p=0}^k (-1)^p \frac{\partial a_{i_0 \dots \widehat{i_p} \dots i_k}}{\partial x_{i_p}} = 0 \text{ for any } 1 \leq i_0 < \cdots < i_k \leq n.$$

The proof of the Poincaré Lemma says that, if the condition is satisfied, then  $\omega = d\varphi$  for

$$\varphi = IH^*\omega = \sum_{i_1 < \cdots < i_k} \left( \int_0^1 t^{k-1} a_{i_1 \dots i_k}(t\vec{x}) dt \right) \sum_{p=1}^k (-1)^{p-1} x_{i_p} dx_{i_1} \wedge \cdots \wedge \widehat{dx_{i_p}} \wedge \cdots \wedge dx_{i_k}.$$

For the special case  $k = 1$ , a 1-form  $\omega = \sum_{i=1}^n a_i dx_i$  is closed if and only if  $\frac{\partial a_i}{\partial x_j} = \frac{\partial a_j}{\partial x_i}$ . When the condition is satisfied, the potential for  $\omega$  is

$$\varphi = \sum_{i=1}^n \left( \int_0^1 t^{1-1} a_i(t\vec{x}) dt \right) x_i = \int_0^1 \sum_{i=1}^n a_i(t\vec{x}) x_i dt.$$

This is the integral of  $\omega$  along the straight line from the origin  $\vec{0}$  to  $\vec{x}$ , and recovers the classical discussion.

**Example 15.5.4.** We claim that  $\mathbb{R}^n - \vec{0}$  is not contractible. The idea is to consider the  $(n-1)$ -form

$$\omega = \frac{1}{\|\vec{x}\|_2^n} \sum_{i=1}^n (-1)^{i-1} x_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n,$$

which gives the higher dimensional version of the winding number (see Exercise 15.45). We have

$$d\omega = \left[ \sum_{i=1}^n \frac{\partial}{\partial x_i} \left( \frac{x_i}{\|\vec{x}\|_2^n} \right) \right] dx_1 \wedge \cdots \wedge dx_n.$$

By

$$\frac{\partial}{\partial x_i} \left( \frac{x_i}{\|\vec{x}\|_2^n} \right) = \frac{1}{\|\vec{x}\|_2^n} - \frac{n}{2} \frac{x_i}{(x_1^2 + \cdots + x_n^2)^{\frac{n}{2}+1}} 2x_i = \frac{1}{\|\vec{x}\|_2^n} - n \frac{x_i^2}{\|\vec{x}\|_2^{n+2}},$$

we get  $d\omega = 0$ .

If  $\mathbb{R}^n - \vec{0}$  is contractible, then by the Poincaré Lemma, we have  $\omega = d\varphi$  and

$$\int_{S^{n-1}} \omega = \int_{S^{n-1}} d\varphi = \int_{\partial S^{n-1}} \varphi = \int_{\emptyset} \varphi = 0.$$

On the other hand, on the unit sphere, we have

$$\omega|_{S^{n-1}} = \sum_{i=1}^n (-1)^{i-1} x_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n.$$

by Example 15.4.2, this is the natural volume form of the sphere, so that its integral cannot be 0. The contradiction shows that  $\mathbb{R}^n - \vec{0}$  is not contractible.

**Exercise 15.47.** Is the potential of a closed  $k$ -form unique? What is the difference between two potentials on a contractible manifold?

**Exercise 15.48.** For a 2-form on  $\mathbb{R}^n$ , find the condition for it to be closed, and then find the formula for the potential of the closed 2-form.

**Exercise 15.49.** Any  $n$ -form on  $\mathbb{R}^n$  is closed. What is its potential?

**Exercise 15.50.** A subset  $U \subset \mathbb{R}^n$  is *star-shaped* if there is  $\vec{x}_0 \in U$ , such that for any  $\vec{x} \in U$ , the straight line segment connecting  $\vec{x}_0$  to  $\vec{x}$  lies in  $U$ . For example, balls and cubes are star shaped. Find the formula for a potential of a closed form on a star shaped open subset of  $\mathbb{R}^n$ .





